

Market for industrial research?

The British government is likely further to damage its support of industrial research by the narrow pursuit of what the markets will bear.

THE British government has found a delightful formula for not spending money on industrial research: if a proposed development promises immediate benefit, let commercial companies provide the necessary funds, but otherwise consign it to oblivion on the grounds that any benefits that accrue will be too far away to matter. Mr Joseph Heller, the author of *Catch 22*, would be proud of the trick, which the British government calls the "near-market" criterion.

The latest application of the doctrine is embodied in the announcement last week (see page 3) that research in the field of thermonuclear fusion at the Culham Laboratory of the UK Atomic Energy Authority will be run down over the next few years. That follows closely on the decision that fast reactor research will be run down, the sale last year of the Plant Breeding Institute to private interests, the decision that the National Engineering Research Laboratory at Glasgow should also be sold to the highest bidder in the next few months, perpetual reluctance to participate in the optional programmes of the European Space Agency and what seems to be the decision, signalled so far by silence, that the programme of support for research and development in information technology, launched with optimistic drumbeats in 1983 under the name of the 'Alvey programme', should be followed by ... nothing. Nobody disputes that public sponsorship of industrial research should be contained within estimates of what can be afforded, and that there should be some calculation of which projects are likely to yield the greatest benefits. There is also ample evidence, going back for decades, that British sponsorship of applied research has been skewed in an imprudent way towards favoured industrial sectors — notably those involved in the manufacture of military equipment (aircraft and electronics). But the now apparently indiscriminate use of the Heller-like formula will do unexpected damage.

The hazards of the policy are linked with the history of the recent past. For most of this decade, academic laboratories in universities and research institutes operated directly by the research councils have suffered most from the British government's attitude towards research. By keeping the available funds constant (after allowing for inflation), but requiring extra from the pattern of spending in civil research, the government has succeeded chiefly in undermining morale. In fields such as agriculture, where there have been unpredicted and real reductions of funds, the damage has been particularly severe.

Consequences

The consequences are now becoming plain. The supply of young people with research experience ready to continue in industry has shrunk, well-qualified graduates from British university courses are too few to satisfy the immediate needs of successful industries and there has been a further decline of the enthusiasm of school-leavers to commit themselves to a training in science. None of that was planned, but it could have been predicted. Young people are not fools, and prefer to embark on careers in fields in which there are jobs available, not in fields in which able people are being put out to grass before their working lives are over. The underlying miscalculation by the British government's

cash-conscious managers was that the sole input to the research enterprise is money and that the sole output of significance is qualitatively the same. One obvious danger now is that the application of the near-market test to applied research will further accentuate the British tendency to opt out of research. Another is that British industry will be technically ill-prepared for what lies ahead.

Fusion

The decision to end support for research in thermonuclear fusion nicely illustrates how the British government applies the near-market test with all the appearance of rationality, but with potentially disastrous consequences. The case for abandonment is strong. Everybody agrees that the practical application of thermonuclear fusion in the generation of electricity is several decades away. The best that can happen is that there will be full-scale machines in operation by the end of the century and that, after a couple of decades of technical development concentrated on the construction of containment walls, heat transfer and the like, people may be in a position to make sensible estimates of the economics of thermonuclear plants 30 years or so from now. It is entirely on the cards that the weight of engineering opinion then will be that thermonuclear power is not an economically favourable way of generating electricity. So, on this occasion, the British government does not expect to shed the development cost to some industrial organization, but equally considers that it should not meet the speculative cost itself. The case for opting out seems all the stronger because Britain continues to contribute to the European programme of fusion research. So why not just save the £10 million a year or thereabouts?

So far, so good. But casual observation of the strategy being followed by comparable countries elsewhere suggests that there may be something missing in the logic. In Europe, West Germany and France (and even Switzerland) continue to spend substantial sums of money on large-scale plasma physics, as do the United States, the Soviet Union and Japan. For better or worse, studies of the behaviour of hot dense plasmas and their stability have become part of the general currency of research. Moreover, a sizeable part of the British research enterprise has grown up in this same field and, conscious of the need that such long-term speculative projects should be made to yield whatever immediate benefits there may be, has been remarkably productive of usable technical innovations in the fields of plasma diagnostics and power engineering. What is now proposed will in due course cause the British to withdraw; even participation in the European programme will be hampered by people's lack of experience in the field. In strictly commercial terms, the gamble is probably justifiable. But there will be even more gloom about the prospects for British science among the generation of young people in whose hands responsibility will rest when thermonuclear fusion is a reality.

That is the most serious of the uncovenanted consequences of the near-market policy, although the further dispiriting of the generation still at work is not an entirely negligible consideration. In this as in other fields, the present British government is skilled at being decisive on narrow issues, but incapable of



appreciating the need to exert an influence on wider matters — the health and adventurousness of the research community, for example. It is like a person skilled at arithmetic but ignorant of mathematics. Sadly for Britain, there is now no prospect that these wider issues will be taken up by the government's new advisory machinery in the field; the Cabinet Office, the chief instigator of the near-market policy, is also firmly in control of the agenda of contemporary discussion. □

Powerless president

Next week's US election will bring long-standing money problems to a head.

IT will probably not matter much to the scientific enterprise in the United States who wins next week's presidential election (which is not to say that researchers are without personal preferences), but it will matter a great deal how whoever is elected spends his time in the next few weeks. For, luckily, the US constitution provides for more than two months between the election and inauguration day. On this occasion more than on most others, this breathing spell will be especially valuable. During the election campaign, both principal candidates for the presidency have been careful to avoid discussion of the most important issue that will confront one of them next year, that of how to deal with the money problems of the United States. So far as the record goes, Mr George Bush has merely said that he will not raise taxes, while Mr Michael Dukakis has let some people believe that he is a protectionist (which may not be true). If either has a strategy for dealing with the problem, he has not disclosed it. Most probably, neither has. Yet the magnitude of the problem increases every day. Two months of reading and thinking will hardly be enough.

For obvious reasons, people working in the United States, still one of the richest economies in the world and certainly the most flexible, do not like to be told that they are running frightening financial risks, but that is the truth awaiting the next president. And while it may be true that the most serious losers would be those who depend most directly on the welfare of the US economy, who may on those grounds consider that their potential misfortunes are strictly domestic, the plain truth is that the US economy is still so large that everybody else will be seriously affected if it should lapse into recession, or worse. It goes without saying that one of the chief casualties of what may lie ahead will be the research enterprise itself.

Here is the problem. First, there is a federal budget deficit, now estimated to amount to \$155,000 million in the financial year that has just begun, but likely to be greater by the end of it. Second, there is the trade deficit, comparable in amount, which remains obdurately insensitive to the decline in the value of the US dollar by about 15 per cent since the Wall Street crash a year ago. Third, there is the way in which the US banking system has been weakened by injudicious lending. The way in which the large commercial banks have been undermined by loans to developing countries, unlikely now to be repaid, may be less serious than the difficulties of the small savings and loans institutions, which are reckoned to require an injection of \$50,000 million of taxpayers' funds if the federal government's promise to safeguard depositors (mostly also taxpayers) is to be kept. Finally, there is the growing indebtedness of those US corporations which have managed, over the past few years, to elude purchase by companies elsewhere.

The reason why these problems will be an immediate crisis for the next president is simply that the indulgence of overseas lenders for the continuation of the federal budget and trade deficits stems from a recognition that nothing can happen while the election is under way; currency traders the world over are too familiar with the US electoral process to sell dollars short at this stage, but they will do so quickly once President Ronald Reagan has handed over, and if his successor should not quickly

spell out a strategy for the future. The trouble is that some obvious strategies are now ruled out. Raising taxes to cure the deficit, let alone to raise funds to keep the savings and loans institutions afloat, may threaten a recession and thus impede the servicing of the debts that US companies have accumulated, most conspicuously in the recent spate of reconstructions and buy-outs financed by the ocean of junk bonds in which they now float. That could precipitate calamity. The only alternative is, sadly, inflation.

Research is likely to suffer either way. By making clamant demands on revenues, corporate debt squeezes out what might otherwise be spent on research and development. Inflation, especially when coupled with high interest rates, works in the same direction. Whatever pre-inauguration reading finds its way onto the new president's desk, the outlook is not cheerful. □

Soviet free science

Soviet science is tackling its problems realistically, but needs luck if it is to solve them.

THE Soviet Union seems to have gone a long way to acknowledge what needs to be done to restore Soviet science to health. That much should be apparent from the report on page 9 of this issue (see also page 3). The election last month of Andrei Sakharov to the praesidium (council) of the Soviet Academy of Sciences was understandably the chief reason why the academy's extraordinary general meeting was a sensation, but it is clear that much else has now crystallized in people's minds. On the harmful consequences of two particular difficulties — the general shortage of first-class equipment and the immobility within Soviet research partly caused by the housing shortage (tradition and mere inertia are other causes) — the academy's managers seem agreed. The absorbing question now is whether their new understanding will enable them to surmount them, even in Mr Mikhail Gorbachev's Soviet Union.

Like Gorbachev himself, the academy needs a little luck, perhaps even a lot of it. For most of the past 70 years, Soviet governments have been generous in their support of science and full of expectations of the benefits that would flow from the energetic application of research. But the same governments have also clapped on the shackles that have made Soviet science less productive than it might have been. Lysenko's genetics is only the most vivid illustration of the way in which the path of inquiry has been guided from above, making people whose instinct and professional need is for free discussion bite off their tongues in silence. For researchers, *glasnost* is not a luxury but a necessity, which is why so many people in the Soviet Union have been heartened by the events of the past two years. Sakharov's election may be symbolic, but it is a powerful symbol. Who can suppose that the academy will act repressively, or be itself repressed, while he remains in office?

But where will the luck come from? The obvious difficulty is that the academy and its institutes cannot opt out of the problems that constitute the daily privations of Soviet life. The lack of housing may make a potentially mobile profession static, but everybody is similarly affected and there is no obvious reason why science should be privileged. Competitive grants should make a difference to the temper of research by helping to distinguish successful projects, but it will take time to make sure that the competitions are running in a way that commands confidence. The choice of the instrument industry as a spearhead of competitive co-existence with the West is wise, but that too will take time to succeed. That is why the academy should now be scouring its research portfolio to identify projects whose instigators can collaborate with groups in the West, preferably on equal terms. For, while Western governments necessarily respond cautiously to the question "What can you do for *perestroika*?", the self-interest of researchers should equally move them in that direction. □

All collaborators held to be responsible for errors

- Inadequate controls led to false conclusions
- Call for guidelines on joint research

Berkeley

WHAT is the responsibility of scientists for factual errors made by their co-authors? That question was addressed by a Stanford University investigation that revealed the improper use of senile patients as 'normal' controls in several psychiatric studies conducted at Stanford's Mental Health Clinical Research Center (MHCRC). A report issued last week by Stanford provost James Rosse blames not only the clinical investigators who chose patients for the study, but also their co-authors in Stanford's department of psychiatry, who performed biochemical analysis on fluid samples taken from the patients.

The report was the culmination of a lengthy university review of the research of Philip Berger, who resigned as director of the MHCRC in 1987 after a university audit found that he had inappropriately spent \$128,000 of research funds from the National Institute of Mental Health (NIMH). At the request of the NIMH, Stanford initiated a review of Berger's research. The reviewers found that three studies of neurotransmitter metabolites in the cerebrospinal fluid of mentally ill patients had included senile patients among their 'normal' controls. Stanford's Ethical Scientific Performance committee judged the use of such controls "inappropriate and possibly confounding". It found that 10 research articles published over the past nine years contained analyses based on the controls, and that in three cases exclusion of the questionable control values changed or invalidated the conclusions of the article. At the

provost's request, the authors of all 10 articles wrote letters of correction or retraction to the journals in which the articles appeared.

The report concluded that the choice of inappropriate controls did not result from a "knowing, purposeful plan to mischaracterize the data", and therefore did not constitute scientific fraud. Because the data were collected 10 years ago, the committee said it was unable to determine whether the choice of the controls resulted from a clerical error that went undetected because of inadequate supervision by senior members of the research groups, or from a conscious decision. The latter, according to the committee, would have constituted a "significant error of judgement". Whatever the specific circumstances, the report suggests that closer supervision of laboratory groups and review of data by senior investigators are necessary to guard against such errors.

Jack Barchas, a professor of psychiatry at Stanford who is a co-author of many of the articles, criticized the committee for releasing its results before those of a current NIMH investigation, and for making "sweeping conclusions" that hold each participant in a collaborative research project "strictly liable" for errors made by other participants. In a statement released with the report, Barchas declines to accept the choice of controls as erroneous, but he points out that the analysis of samples carried out by members of his research group was beyond reproach.

"In an ideal world it is desirable that all collaborators on a given project should have equivalent knowledge concerning each other's databases", says Barchas, "but that view does not reflect the reality of the research process when two or more independent units interact in the service of dealing scientifically with a problem that neither group can approach alone." Barchas warns that a policy of holding every author responsible for the errors of a colleague will make interdisciplinary collaborations impossible.

In his report, Rosse notes that there are no generally accepted standards for the degree of responsibility borne by each participant in a collaborative project. He has asked Stanford's faculty senate to formulate guidelines for interdisciplinary collaboration that would outline the responsibility of co-authors and include procedures for data review and training of research teams.

Marcia Barinaga

Fusion research ends

London

THE reluctance of the British government to fund research with long-term applications led to the announcement last week of the winding down of fusion research. After the cuts made in fast reactor research in July (see *Nature* 334, 278; 1988), these latest cuts do not come as a surprise.

Britain's fusion research is based at the United Kingdom Atomic Energy Authority's Culham Laboratory where the Joint European Torus (JET) is sited. The government has agreed to fund JET until the scientific programme is completed in 1992, and will continue to fund research on the tokamak, a toroidal magnet for containing plasma. But it is winding down funds for other fusion research. £5 million will be cut from the annual budget of £21 million over the next three years. The first to go may be research on a reverse field pinch, a smaller version of the tokamak, and a range of studies, for example on environmental safety, materials and work for the Next European Torus (NET).

The future of 150 staff is now under threat, although the AEA will try to transfer the staff to commercial work. The Culham laboratory has doubled its income from outside contracts to about £8 million. A spokesman for the laboratory said work there would continue as long as there is a programme of fusion research in Europe.

Christine McGourty

Early launch for Hubble telescope

Washington

A CHANGE of plan by the Pentagon will allow a long-awaited scientific mission to get into orbit nearly two months ahead of schedule. The Hubble Space Telescope is now set for a December 1989 launch aboard the US space shuttle, a time slot originally reserved for the Department of Defense. But Defense officials have asked the National Aeronautics and Space Administration (NASA) for a delay. Worried that a one-month delay would move the two missions too close together, NASA proposed that the classified military mission swap launch dates with the space telescope, and the Pentagon agreed.

NASA officials do not expect any problems in getting the space telescope ready for the earlier launch date. The spacecraft is being stored in northern California, where it will undergo comprehensive ground tests later this month.

The only aircraft capable of transporting the 24,000 pound spacecraft is the Air Force C-5A transport, and NASA is building a special carrying case to allow air transport. If a C-5A is not available, the telescope will be transported by sea.

Joseph Palca

Moscow News

NATURE has made an arrangement with the *Novosti* Press Agency for gathering regular news from the Soviet Union. First, *Nature* has established a Moscow Advisory Group consisting of Academician Roald Sagdeev (chairman), Academician Vitaly Ginsburg (Lebedev Institute) and Professor Maxim Frank-Kamenetskii (Institute of Molecular Genetics) which will, among other things, identify to the editor of *Nature* and to *Novosti* topics to be covered. The group will also provide help and guidance to the reporters assigned to these tasks as required. The first contribution appears on page 8 of this issue. □

Volte-face on controversial French abortion pill

Paris

A RAPID about-turn last week saw the controversial abortion-inducing pill RU 486 taken off the market by its manufacturers, Roussel-Uclaf Laboratories, only to be put back again 48 hours later under pressure from the French Health Minister, Claude Evin. In a ministry communiqué, Evin said that he was acting "in the interest of public health", to uphold women's rights to voluntary abortion under a 1975 statute. He warned Roussel-Uclaf that a 1968 law gave him powers, through the Industry Ministry, to seize the patent for RU 486 (like any other pharmaceutical product judged "essential for public health") and to hand it to another laboratory if the company did not resume production.

Roussel-Uclaf Laboratories decided on 26 October not to go ahead with plans to market RU 486 because of the "emotional reaction of a sector of public opinion in

France and abroad" provoked by news of the drug. These critics say that a black market for the drug will emerge, leading to a trivialization of abortion. The company's decision came just one month after RU 486 was given government approval in France and China for administration under medical supervision (see *Nature* 335, 486; 1988). It now turns out that anonymous anti-abortionists have threatened company employees and their families, an action which Evin condemned as "scandalous" and "cowardly".

When news of the withdrawal reached a meeting of 9,500 doctors, including Emile-Etienne Baulieu who first synthesized RU 486, at the World Congress of Gynaecology and Obstetrics in Rio de Janeiro, there were outraged protests. This response is believed to have encouraged Evin to intervene last Friday. Speaking on French television, he said that "this product does not belong to a laboratory, it

belongs to women".

RU 486 blocks the action of the hormone progesterone, essential at all stages of pregnancy, and has been shown in clinical trials to be safe and reliable in more than 95 per cent of cases, when used with prostaglandins. The French Ministry of Health approval limits distribution of RU 486 to the 800 registered abortion clinics where it is to be given by a doctor only in association with prostaglandins and only within the first 49 days of pregnancy.

The principal advantage of the drug is that it allows some women the option to terminate an unwanted pregnancy without the need for anaesthetic or surgery. But around one in five women could experience excessive bleeding and there is a risk that the fetus would not be expelled. In such cases, women are required, by a Health Ministry stipulation, to undergo surgical abortion, as the surviving fetus could be deformed.

Following approval in France and China, Roussel-Uclaf planned worldwide commercialization of the drug, and applications have already been filed in Britain, the Netherlands, Sweden and Spain. But threats of a campaign by the US Right to Life Committee to boycott Roussel-Uclaf products and those of majority shareholders, Hoechst, together with a strong Catholic anti-abortion lobby in France, led the company's directors to reconsider at an emergency board meeting last week. The strength of conservative Catholic opinion in France has recently made alarming headlines, after cinemas showing Martin Scorsese's film, *The Last Temptation of Christ*, were subjected to arson and tear-gas attacks.

Restrictive provisos in the French government's approval meant that commercial returns from RU 486 would in any case have been less than Roussel-Uclaf had hoped if overseas markets could not be exploited. Only half of French women seeking abortions would be eligible and the drug's withdrawal is not expected to affect the number of abortions carried out. For the moment, Roussel-Uclaf's about-turn affects only France and no decision has been taken regarding resumption of worldwide distribution, where threats of damage to the company's image through boycotts and adverse publicity remain.

But RU 486 is seen by the World Health Organisation (WHO) as a useful means of birth control to reduce suffering in developing countries. Roussel-Uclaf had already signed a contract to sell WHO the drug at cost price. If necessary, the company says it will give WHO a licence to sell the drug abroad under its own name. Apart from its use as an abortifacient, RU 486 has potential therapeutic value in treating cancers of the prostate, a less controversial prospect that researchers are following up.

Peter Coles

Genentech loses appeal against Wellcome

London

GENENTECH has once again lost the fight in Britain over the clot-dissolving agent tissue plasminogen activator (TPA). The British Court of Appeal this week upheld the ruling of the High Court in favour of the British drugs company Wellcome which is challenging the patent held by the US company. Wellcome will greet the decision with satisfaction, but it deals a severe blow to Genentech, coming at a time when sales of TPA are falling short of predictions and questions are being raised over the company's promotion and testing of the drug (see *Nature* 335, 751; 1988).

This week's decision comes more than a year after the High Court ruling which both sides regarded as a victory. Wellcome said it vindicated its arguments that Genentech's patents failed to meet the criteria of novelty and inventiveness but Genentech lawyer Simon Cooke disagreed, saying that the decision showed that the judge believed the product was novel and inventive.

After Genentech's appeal was dismissed this week, Cooke said that Genentech would decide within the next few weeks whether to appeal to the House of Lords; the court of appeal granted leave for the company to do so. An appeal would have to be filed before the end of January 1989 and would probably be held next summer. Cooke said he could not comment on the ruling before studying the 317 pages of judgement.

After the initial ruling in the High Court, Cooke said the judge had "misunderstood the technology", and had made a decision

that was "fundamentally wrong".

But the Court of Appeal's dismissal of Genentech's claims comes as no surprise to Britain's patent lawyers. "The case was very simple and the court made the right decision", says patent lawyer Iain Baillie. He said Genentech was trying to stop other researchers making the same material through other techniques. The company was being "a little too greedy", and if it appeals to the House of Lords that appeal is unlikely to succeed, he said.

Christine McGourty

Mexican reactor goes on line

Boston

AFTER two decades of planning and construction, Mexico's first commercial nuclear power plant came on line last month after the Mexican government gave operators permission to activate one of the two 654-megawatt nuclear reactors in Laguna Verde. The plant, near the Gulf of Mexico in the state of Veracruz, has been completed for months, but it has been the focus of nationwide controversy. Opponents claim that the government waited until after local elections to start up the plant in order to subdue opposition to the reactor.

The opening of the Laguna Verde plant is said to have been met with protests from a coalition of environmental groups, including a permanent vigil outside the reactor site.

Seth Shulman

Small university departments threatened with closure

London

THE reorganization of teaching and research in UK universities, set in motion principally by restrictions on resources, shook the Earth sciences community earlier this year. Now it is the turn of chemistry and physics.

The two committees set up by the University Grants Committee (UGC) to plan the restructuring published their reports this week and both advocate concentration of available resources in fewer departments. But both also lay down strict guidelines to determine which departments should survive: those with at least 20 full-time staff and 200 full-time students. Departments below this minimum level should close, merge with other departments or expand. There will probably be many interdepartmental mergers and some closures. At present there are 50 departments of chemistry and of physics; after restructuring there may be only 30 of each.

That size should be the key factor in determining the future of departments has raised the hackles of chemists and physicists in small departments. "Big is not necessarily beautiful" has been the common cry. And even those in large departments, the 'winners' according to the new reports, say size should not be the sole criterion.

Research income is one standard used to measure the success of a department. Statistics from the Committee of Vice-Chancellors and Principals show that the chemistry department of the University of Essex attracts more income than any other chemistry department in Britain, but it has only 130 full-time students. The chemistry department of the University of Bristol has three times as many students and only two-thirds the amount of income from research.

The response of the science community to the reports has been to urge the UGC not to implement the recommendations too rigidly. The physics committee acknowledges the benefits of small departments: they do come up with "bright ideas", and have good staff-student relations. The chemistry committee is more reluctant in recommending the closure of small departments; an abrupt end is better than the present "demoralizing piecemeal attrition" due to lack of resources, it says. But among the benefits of having fewer and larger departments is their enhanced visibility and the increased probability of achieving international recognition. Both committees urge that every university should maintain some presence in each subject.

If the restructuring of other subjects

proceeds according to size, the future of entire universities could be in doubt. Smaller universities tend to have smaller departments in several subjects. If government policy does not favour the closure of individual universities, this leaves a seemingly insoluble problem, says Professor Sam Edwards of the University of Cambridge, chairman of the physics review committee. Sir Peter Swinnerton-Dyer, chairman of the UGC, says that the future of chemistry and physics departments will not be planned without taking account of the effects on the university as a whole.

As well as advocating concentration of resources, which was bound to upset those to lose out, the committees tackle the most serious problems now facing academic science — lack of equipment and difficulty in attracting students and young academic staff. The committees both ask

for £30 million immediately for new equipment.

The chemistry report recommends that students be given greater opportunities to work in industry during vacations and that postgraduate training should include formal instruction and should not be slanted towards the needs of contemporary industry. It also says that basic research is in "serious danger" as resources are increasingly diverted into strategic and interdisciplinary research, and urges that it be protected.

The physics committee says that teaching should be more comprehensible to students and calls for a review of the difficulty of examinations, and for more high-quality applied physics degree courses. Many academic physicists are depressed at what they see as a decline in physics research in Britain. An international review carried out for the committee by a policy studies group of the Royal Society shows that Britain's share of publications in physics declined from 7.7 per cent in 1973 to 5.7 per cent in 1984.

Christine McGourty

Soviet shuttle fails at last minute

Washington

THE first flight of the Soviet space shuttle came within 51 seconds of lift-off on Saturday, 29 October, but a malfunction on the Baikonur launch pad caused a computer to abort the countdown.



51 seconds to go — and Buran is grounded.

The flight has now been postponed indefinitely.

The setback came when the rocket was being freed from the servicing towers and the engines were being prepared for firing. The platform which carries the system that sets the rocket's gyroscopes failed to separate from the rocket to a safe

distance. According to the Tass news agency, the fuel is being drained from the rocket and the engineers will then try to determine the cause of the problem.

The Soviet version of the space shuttle — called *Buran*, meaning snowstorm — bears a striking resemblance to its US counterpart. It has the same delta-wing design for the orbiter and strap-on boosters, and it is attached to a large, external fuel tank. But many of the similarities are superficial, according to John Pike of the Federation of American Scientists, who has been studying the Soviet space programme. For example, the four booster rockets on the Soviet shuttle use a liquid rather than a solid propellant, and the main engines are integral with the fuel tank and are discarded after each mission.

The orbiter itself is said to have three engines for manoeuvring while in orbit, but these do not appear to be as large as those used by the US space shuttle orbiter. *Buran* is said to be able to carry a crew of seven, but the crew quarters appear to be more cramped aboard the Soviet shuttle, suggesting that it is to be used more as a ferry vehicle between Earth and an orbiting space station, rather than a stand-alone platform for manned space activities.

The payload capacity for the Soviet shuttle is around 60,000 pounds, a little less than the current US shuttle lift capacity. But the Energia booster system can be used to lift unmanned payloads as well, making it a more flexible vehicle.

The first mission was intended to be unmanned, completing two orbits of the Earth.

Joseph Palca

Referendum on Massachusetts nuclear power

Boston

THE future of nuclear power in Massachusetts will be determined in November by a state ballot on a measure that would permanently close the state's two nuclear power plants. If the measure is passed, Massachusetts would become the first state to outlaw commercial nuclear power for the generation of electricity.

Proponents of the nuclear shutdown have capitalized on the woes of the state's Pilgrim nuclear reactor to drive home their message. The Pilgrim plant has been shut for over two years because of safety violations, and is rated by the Nuclear Regulatory Commission as one of the least safe and worst-run nuclear plants in the country. They also note that Yankee Rowe in western Massachusetts, the state's only other nuclear power plant, is the oldest reactor in the United States and lacks the concrete containment shell found in newer reactors.

Opponents of the measure have stressed the economic costs of an abrupt shutdown of nuclear power production in the state, both in replacement energy from imported fossil fuels and in compensation to the plant's owners. Opponents of the shutdown have also noted their objection to what some call the "scare tactics" used by the measure's supporters in their depiction of the potential dangers of nuclear power. The group has organized an impressive collection of scientists and engineers, including several Nobel laureates and university presidents, to plead the case of the opposition in public forums and on television commercials.

The most recent polls indicate that the measure may lack the majority of voters needed to carry it, but observers say that the race will be a close one. A similar measure in Sacramento, California, seeking to close the Rancho Seco nuclear plant last June, lost by only 2,200 votes — less than one per cent of the vote. Local utilities have raised a staggering \$7.3 million to fight the measure, more than twenty times the amount raised by the bill's supporters.

Kent Hansen, professor of nuclear engineering and associate director of the Energy Laboratory at Massachusetts Institute of Technology, is a vocal opponent of the shutdown. He does not see the measure as a solution to the state's energy problems and is frustrated by what he feels are "misleading" statements made by the measure's supporters. "I would hate to see a decision about what we want to do about nuclear power be based on misinformation", he says. **Seth Shulman**

New TV satellite is launched — but will anyone watch?

Paris

THE French direct-broadcast television satellite, TDF-1, which was launched successfully on 28 October by an Ariane 2 rocket, puts an end to a ten-year telecommunications squabble. But with only two subscribers for its five channels, it may end up an expensive white elephant.

TDF-1 was conceived in 1977 as the means by which France — and Europe as a whole — could keep pace in the international television market in the face of overwhelming dominance by Japan. The satellite will be able to transmit high-definition television (HDTV) pictures with digital stereo sound to homes throughout most of Western Europe, provided they have access to the right kind of parabolic antenna. The new broadcast standard, known as D2 Mac Packet, which TDF-1 uses, was finally adopted at a meeting in Dubrovnik in May 1986 in order to give Europe a single norm. At present, French televisions use the SECAM norm, while much of Europe has opted for PAL. Japan and the United States use a totally different standard, called NTSC. This chaotic incompatibility has proved a headache for manufacturers.

Ever since the West German and French governments agreed in 1979 to build the same satellite, the career of TDF-1 from drawing-board to geostationary orbit has been fraught with difficulties. In 1984, the director of the French telecommunications network (DGT), Gérard Théry, pronounced TDF-1 already out of date. DGT wanted its own, more modern

satellite, Telecom 1, to be adopted instead. But the largely Franco-German consortium, Eurosatellite, which built TDF-1 and its West German sister, TVSAT1, has managed to persuade successive governments to back its project, despite a doubling of costs to around FF1800 million (\$290 million) in ten years.

In May 1986 Eurosatellite was set back when a third-stage rocket motor failure halted Ariane launches until November last year. A cheaper, 17-channel satellite, ASTRA, being built by Luxembourg (due for launch next month) had meanwhile come on the scene, and TVSAT1 failed when its solar panels failed to open despite a perfect launch at the end of last year. Originally scheduled for March, the TDF-1 launch was put back after Jacques Chirac offered to India the satellite's earlier place aboard an Ariane rocket.

Now that TDF-1 is finally in orbit, the question of who will use it remains. So far, only the West German Bundespost and a new seventh French channel, *La SEPT*, have signed contracts. With subscriptions to the French cable network disappointingly low — only 50,000 subscribers for the million homes so far equipped with cable — all the other French television stations have declined to take up FF40 million-a-year. But the French hope to launch TDF-2 in September 1989, if a FF1000 million shortfall can be made up from the private sector, while West Germany has booked space on Ariane in February 1990 for its TVSAT2.

Peter Coles

Union upheld against frivolous challenge

Boston

HARVARD University last week lost its appeal to overturn a plebiscite that would allow the nearly 4,000 clerical and technical workers at the university to join a union. An administrative law judge handed down a 40-page ruling that dismissed each of Harvard's claims that the election was held improperly.

The decision is a resounding victory for the union, an affiliate of AFSCME (American Federation of State, County and Municipal Employees) in a long and often bitter labour dispute with the university (see *Nature* 335, 389; 1988). The contest has been watched closely by labour leaders around the nation. AFSCME officials have gone so far as to call the election a "test case" for the labour movement in the United States.

In the recent ruling, the judge not only dismissed Harvard's claims, but labelled some of their challenges "frivolous". The

judge wrote that to his knowledge no case had ever been brought claiming misconduct while presenting evidence "devoid of coercion, misstatements, bribery, violence, or vandalism". In response to Harvard's contention that election-day processes had effectively denied some 1,700 members of the electorate equal access to the polls, the judge quoted from an earlier ruling stating "puffs of this sort of smoke do not indicate the existence of any flame".

The university now has a fortnight to appeal against the ruling. But because of the forcefulness of the decision, union supporters say they will press Harvard to accept and to give up what many now see as a hopeless effort to block union certification.

University officials say they are currently in the process of "reviewing the decision in depth" and have not decided whether to seek an appeal on the ruling.

Seth Shulman

French city heading towards Technopolis

Paris

THE first in-depth review of an experiment in technological innovation introduced in France in 1984 has recently been published* by a group of economists at the Centre National de la Recherche Scientifique (CNRS). The project aims to stimulate regional nuclei of high technology competence (*technopôles*) by creating a network of consultants from local universities and state research centres who collaborate with industry. The experience at one of these *technopôles*, Sophia-Antipolis in the South of France, suggests that a new phenomenon is being born.

Instead of the traditional science park, which can be no more than a juxtaposition of research and industrial competence, Sophia-Antipolis could, the report says, become a 'technopolis' — a flexible infrastructure which promotes technology innovation as well as technology transfer. Such transfer is necessary, but is not sufficient for manufacturers to respond to demands for creation of new technology.

Companies at Sophia-Antipolis, which include the information technology multinational Digital Equipment and the electronics company Thomson-Sintra, are moving away from rigid management and production structures which only exploit existing technology developments. By using modular, multi-purpose production lines and creating project-oriented 'task forces', including consultants from universities, companies can create new technologies which are tailored to the client's demands. In this way, Thomson-Sintra was able to use expertise gained in the development of sonar devices for submarines to create equipment to locate faults in the cooling system of the prototype Superphénix nuclear energy reactor, which has been out of service for over a year.

The CNRS report finds few faults with developments at Sophia-Antipolis. Rather, it is concerned that *ad hoc* principles of collaboration be reinforced in order to assure a "perennial success" in responding to demands for new technology. The role of industry in researcher training is seen as paramount — a role which successive governments have criticized as insufficient. But leading engineering academics also need to review their role. At present, few offer facilities for research to doctoral level and graduates tend to shy away from careers in state research organizations, such as CNRS. Without better-trained research engineers, France could fall behind in the all-important race for new technologies, despite successes like Sophia-Antipolis.

Peter Coles

**Technopôle et Développement*, available from J.-L. Ravix, CNRS, Avenue Albert Einstein, 06560 Valbonne, France.

New initiative in Japan to develop neural networks

Tokyo

JAPAN is getting ready to make a big splash in the world of neural network computers. Within the next few months the Ministry of International Trade and Industry (MITI) will announce the formation of a study group to lay plans for a new national neurocomputer project. Although details are not known, it is expected that the project will match the fifth generation computer project in scale.

Japan has solid foundations in neurocomputing set in the 1960s when there was a feverish burst of research activity in laboratories around the world. When the field went out of fashion, a few basic researchers continued working in Japanese laboratories. Now that the topic is generating excitement worldwide, their research and new projects could provide the core for a national effort.

Nobuyuki Ohtsu, section chief of the mathematical engineering section of the computer science division at MITI's Electrotechnical Laboratory (ETL) and a key adviser to MITI, expects that the project will have four main topics: theory and fundamental research on neural network models; architecture and parallel processing; devices, including optical devices; and application technology. But he stresses that the topics are still "not so clear".

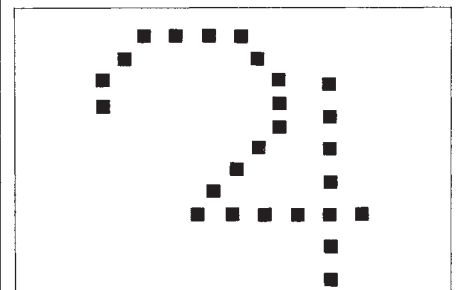
Like Ohtsu, Professor Shunichi Amari of Tokyo University hopes that the project will concentrate on theoretical work in the initial stages. Amari is another key MITI adviser and one of those who continued research during the 1970s. He says much more needs to be done before a commercial neurocomputer can be built, and at this stage it is impossible to predict what kind of machine it will be.

The study group, which will be established next spring with a budget of ¥20 million will look at optical, fuzzy and parallel computing as well as neurocomputing. But the chances are that it will come up with a national project that will attempt to replace the fifth-generation computer project while avoiding its shortcomings.

Despite all the publicity — and ¥30,000 million in government support since 1982 — the fifth generation computer project is not seen as a success. Industry will be reluctant to contribute its best researchers to a second project based around a single institute. And ETL researchers are wary of getting dragged into management of another large-scale national project. Thus it is expected to take MITI one or two years to reach a final consensus on how the neurocomputing project should be structured.

Gen Matsumoto of MITI's Electro-technical Laboratory in Tsukuba, who is studying the squid giant axon, heads the 'bioarchitecture' group within a MITI-sponsored 'bioelectronics devices' project, now in its second year, aimed at developing new algorithms and simulations for neurocomputing through the study of living organisms. Industrial partners include Fujitsu and NEC.

Before the national project takes off, substantial government funds will be



What numbers are superimposed in this figure? When asked this question a neural network model developed by Kunihiro Fukushima separates out first 4 then 2 as expected when simulated on digital computer . . . it then goes on to prove its 'intelligence' and extracts 1.

poured into neurocomputing at a new institute set up by the Ministry of Post and Telecommunications. The Advanced Telecommunications Research Institute (ATR) in Osaka was established in 1986 with funds from private industry and the Japan Key Technology Centre. ATR has an annual research and development budget of about ¥9,000 million (\$70 million) and 200 researchers, about ten per cent of whom are working on neural network projects.

For example, Mitsuo Kawato of ATR's Auditory and Visual Perception Laboratory is collaborating with Bridgestone in the development of neural networks to control robot movement.

And Sei Miyake of the same laboratory is developing the 'neocognitron', a neural network model developed by Kunihiro Fukushima of NHK Science and Technical Research Laboratory which can distinguish and read handwritten numbers (see figure). Miyake has adapted the 'neocognitron' to run on a \$2 million NCUBE parallel-processing computer and it can separate out and 'recognize' 4-digit numbers in 0.8 seconds.

Although MITI and the Ministry of Post and Telecommunications are arch rivals, Kawato says there are no barriers between researchers in the field of neurocomputing in Japan. Nearly all the researchers at ATR are seconded from private companies.

David Swinbanks

US sea lions join casualties

Berkeley

It has been a bad year for seals and sea lions worldwide. While Baltic and North Sea seals are dying of viral infection (see *Nature* 355, 3; 1988), sea lions on the California coast are falling victim to two separate maladies, and as in the North Sea case, questions have been raised as to whether pollution is to blame.

Since June, over 100 sea lions suffering from leptospirosis have been taken to the California Marine Mammal Center outside San Francisco, and half have died. Leptospirosis is a bacterial disease that causes kidney damage and infects cattle, swine and rodents, as well as humans. It first appeared in California sea lions in 1970. Neylan Vedros, of the University of California at Berkeley School of Public Health, traced the 1970 outbreak to the rodent population in island rookeries where the sea lions raise their young. He says the disease is probably continually present in the sea lion population at sub-clinical levels, causing periodic outbreaks when population immunity drops sufficiently. A 1984 outbreak killed about a tenth of the population, and some environmentalists question whether this year's outbreak, because it follows closely on the heels of the last, indicates a pollution-induced weakening of the animals' immune defences. Vedros, an immunologist, says there is no evidence of that.

In the past few months, a second illness characterized by severe seizures has appeared in sea lions off the central California coast, near San Luis Obispo. Mary Jane Schramm, of the California Marine Mammal Center, says 28 ailing sea lions and two fur seals have been brought to the centre, and eight have died. Because the outbreak has been localized to a region of the coast where there is oil exploration, in addition to other industry, the centre has asked for help from the US Environmental Protection Agency (EPA) in investigating whether a pollutant may be to blame. The symptoms are consistent with heavy metal poisoning, says Vedros, who is testing animals for heavy metals as well as for signs of bacterial or viral infection.

Peigin Barrett, director of the Marine Mammal Center, says that several years ago there was a similar seizure illness affecting male sea lions off the northern California coast, but an investigation by EPA specialists and the Marine Mammal Center could not find its cause.

Because of their position at the top of the food chain, seals and sea lions may be sensitive indicators of toxin accumulation in the fish they eat, and environmental groups are poised to point the finger at pollution.

Marcia Barinaga

Paris AIDS retrovirus conference

Paris

"THIS is the year of realism" in AIDS research, said Maxime Schwartz at the third 'Cent Gardes' meeting on retroviruses of human AIDS organized by Pasteur Vaccins and the Fondation Mérieux near Paris last week. While a panel of leading researchers painted an image of optimism for the press, albeit qualified by the lack of any real success in the quest for a vaccine, it was clear to most of the 300 invited participants that nothing substantially new had happened since last year.

Jeanne-Marie Leconte, president of Pasteur Vaccins, contrasted the disappointingly slow progress in the search for a vaccine, given the extraordinary initial progress in understanding the human immunodeficiency virus (HIV), with Pasteur's preparation of a rabies vaccine within two years. It took another 80 years before the virus was isolated and a further 24 years before its genome was sequenced.

Not surprisingly, many papers concentrated on the small strategic handhold so far available on HIV — the affinity of the virus envelope glycoprotein gp120 for CD4 receptor sites on some host cell surfaces. But attempts so far to immunize primates with purified gp120 have failed to protect the animals from infection. Attempts to use

soluble CD4 analogues to inhibit viral binding (see *Nature* 334, 557; 1988) are also meeting obstacles. Apart from the very short half-life of CD4 (less than 2 hours in humans, according to Daniel Capon of Genentech), and the ability of HIV mutants to sidestep this receptor, Richard Axel of the Institute of Cancer Research at Columbia University is worried that the immunoglobulin model for CD4 analogues may turn out to be inappropriate. "If this analogy is wrong, then all we think may be wrong", said Axel.

The meeting's organizers deliberately excluded epidemiological studies and also decided not to include reports on public health measures. Human clinical trials of therapeutic agents were also little touched upon. But Donald Abrams, of San Francisco General Hospital, reported on trials using oral dextran sulphate with AIDS patients. Abrams confirmed earlier fears (see *Nature* 334, 369; 1988) that the US Food and Drug Administration's deregulation of unapproved imported drugs with known therapeutic use against AIDS would handicap attempts to carry out controlled clinical trials. Japanese colleagues, he said, were likely to stop collaborating if use of dextran sulphate could no longer be controlled.

Peter Coles

Five years of French AIDS coverage

Paris

NEWSPAPER coverage of AIDS has revolutionized the public image of the scientist, says a report* from the French Centre for Medical Research and Social Sciences (CERMES). But if the press has also helped to transform the disease from 'mystery gay cancer' in 1982 to a 'social phenomenon' today, the authors argue that we can expect a future waning of media interest if the disease is seen as affecting mostly disinherited members of society and populations of poor countries.

The report analyses the coverage of AIDS in six national French daily newspapers, from the first article in January 1982, up until the second international

conference on AIDS in Paris in 1986. "More than on any other occasion", says the report, "during the history of AIDS, research has had news value." Reporting of the dispute between US and French laboratories over patent rights to the AIDS antibody test kit showed that research is more than a matter of intellectual prestige. Political and economic issues are also involved. Researchers were not slow to exploit this advantage, complaining to the press of underfunding.

Because the number of cases is continually growing, statistics in themselves always have news value. But if the press has helped to create notions of risk, global responsibility and the emergence of AIDS as a 'social reality', the report finds that the daily newspapers have also helped to engender fear, albeit while trying to inform. This has been exacerbated by the absence of effective therapies, despite the extraordinarily rapid progress in understanding of the disease.

The report suggests that the present period of intense media interest in AIDS could be followed by "rejection and lack of interest" if the disease is perceived as affecting mostly socially deprived groups and populations of "disenfranchized continents".

Peter Coles

Franco-Australian research agreement

Paris

THE French Research Minister, Hubert Curien, and his Australian counterpart, John Button, last week signed an agreement to cooperate in science and technology. This move marks an attempt to restore political détente between the two countries, at loggerheads over French nuclear testing in the South Pacific as well as over the sinking of the Greenpeace ship, *Rainbow Warrior* by French agents.

Peter Coles

*C. Herzlich and J. Pierret *Une maladie médiatisée. Le SIDA dans six quotidiens français*. CERMES-CNRS/INSERM/EHSS 1988.

General meeting of Soviet academy a sensation

Moscow

LAST month's extraordinary meeting of the Soviet Academy of Sciences (18–20 October) was the most dramatic event of its kind I have had to report in 30 years. Despite the news hunger of the Soviet people engendered by *glasnost*, nobody had expected that the drive for renewal (*perestroika*) would so quickly touch the bulwark of Soviet science, the academy.

During the meeting, and against a background of complaint about serious problems and shortcomings in Soviet science, the academy voted out half its praesidium, but elected Academician Andrei Sakharov to membership, received 500 million rubles of state funds for priority research programmes, overhauled the system of financing research, decided to set up a foreign trade organization and discussed ways of infusing academic research with youthful talent.

The general meeting had been preceded by an unusual event — a poll among the 400 members and corresponding members of the academy seeking opinions of the state of fundamental science in different fields. Academician Guri Marchuk, president of the academy, told journalists that the poll had shown that only about 40 per cent of Soviet research projects were considered to exceed or match world standards — fields such as geophysics and ultra-deep drilling, thermonuclear fusion, laser physics and high-temperature superconductivity. Those polled believe the Soviet Union to be behind in information and computer technology and in genetics (condemned as bourgeois science under Stalin).

Marchuk and others noted that theoretical research is of a high standard, but that the shortage of sophisticated instruments had meant that many outstanding theories had not been quickly confirmed; Soviet leadership in some fields had been lost because the needs of fundamental research had too often been overlooked.

On instruments, the academy says that Soviet industry is not interested in manufacturing sophisticated instruments in small numbers, and that its own manufacturing capacity is as yet too small. Imports of equipment are restricted by the shortage of hard currency and by export restrictions by some Western governments, which have a "negative effect on world science, which cannot use to the full extent its intellectual potential for tackling global problems".

The academy's new trading organization has been prompted by the equipment need. Marchuk said that "we must learn to sell ourselves" and to follow Japan in realizing that "the trade in manufactured goods is more profitable than the trade in licences". Some of the output of the new

instrument plants will be exported, and there are plans for joint ventures.

To stay ahead in areas where Soviet science leads and to catch up in others, the government has allotted an extra 500 million rubles to certain priority research programmes which, unusually, will be chosen competitively. One of 18 such programmes is that in high-temperature superconductivity, directed by a scientific council headed by Marchuk. But Academician Yuri Osipian, a member of the council, says that although the organizations involved have been given everything they need, they are "looking for some instructions from above".

Competition will also be used to determine the allocation of funds in 100 other areas of lower priority while, from next year, all research programmes at academy institutions will be supported in a new way. Funds will be allocated to projects that experts judge promising. Laboratories not receiving funds may have to work for a related industry.

The academy will also receive funds from government ministries and departments to support research programmes commissioned by the State Committee for Science and Technology. Other funds will flow from contract arrangements with organizations and enterprises.

There is also to be, for the first time, a system of competitive research grants, beginning in information and computer science. Academician Yevgeny Velikhov, a vice-president of the academy, said that there had already been 300 applications, and that awards would be decided by a special commission in November, initially for one year but subsequently for three.

Last month's meeting was also occupied with the problems of age and youth. Many complained that the shortage of youthful talent had become worse in recent years, partly because of the lack of funds for the encouragement of young researchers and partly because of the obstacles to the dismissal of incompetent and lazy workers. The lack of housing was recognized to restrict the mobility of scientists.

But most excitement at the meeting was engendered by the elections. Although the practice of electing directors of institutes and heads of departments and laboratories is gaining ground the electoral procedure is far from perfect, and has yet to be thought out. A secret ballot of all the workers in an institution may produce the most acceptable administrative manager rather than the best organizer of research.

The general meeting of the academy, however, rejected the proposal (already imposed on political leaders and industrial managers) that the terms of office of institute directors and other research leaders should be limited to ten years. One

speaker argued that the rule would have banished Pyotr Kapitza, who founded the Moscow Institute of Physical Problems after his return from Cambridge in 1934.

The academy's own elections, brought forward from 1990 by the new rules requiring members of the praesidium to be 70 or younger, caused a sensation in academic circles — indeed, in the Soviet Union as a whole. Five new vice-presidents were elected without difficulty (see *Nature* 335, 753; 1988) as were eight new departmental secretaries nominated by their departments. But there was heated debate about the election of the remaining members of the praesidium.

The election of Andrei Sakharov, in itself remarkable, also caused a stir because he had been nominated by Academician Roald Sagdeev, former director of the Institute of Space Research, who had resigned his post six weeks earlier (see *Nature* 334, 189; 1988). Sagdeev had also been nominated, but stood down in favour of Sakharov, saying that the praesidium should include "people with a broad vision and developed sense of moral and social responsibility to the nation and the world". Marchuk said that Sakharov's election was "further proof of the success of the drive for democratization".

Among other comments on the praesidium elections, physicist Andrei Gaponov-Grekhov (also excluded because only one physicist was eligible) ranked the three candidates in the order "Sakharov, Sagdeev and Gaponov-Grekhov", while mathematician Israil Gelfand said "For the first time, we have had to choose between the good and the best" and went on to plead that all three should be elected. But a formal proposal by Academician Vitaly Ginsburg that members of the praesidium should be elected as scientists, without regard to discipline, was voted down.

The question remains of why the praesidium should be restricted to 47 members, also the reason why sociologist Academician Tatyana Zaslavskaya was not elected last month. In an emotional speech in her support, Academician Natalya Cekhtereva, director of the Institute of Experimental Medicine, said "we have only one woman in the praesidium, and she is in the picture", which was a reference to the portrait of Countess Yekaterina Dashkova, president of the academy in the late eighteenth century.

On another historical theme, the general meeting voted unanimously for the posthumous reinstatement as a full member of Nikolai Bukharin, a close colleague of Lenin who was brought to trial during the Stalin terror, executed and stripped of his academic title at a general meeting of the academy in 1937.

Yuri Kanin
Novosti

Science policy in Australia

SIR—The Australian federal government has placed great emphasis on science and technology to underpin Australia's future economic performance. A recent policy statement on higher education describes the government's agenda for sweeping changes designed to align academic and

Table 1 Career prospects
Perceptions of career prospects in science in Australia as poor

1. Overall	80%
2. Nationality	
Australian	79%
non-Australian	80%
3. Experience (yrs post-PhD)	
<3 yr	82%
3–5 yr	68%
>5 yr	86%
Expectation of tenurable next appointment	
1. Overall	15%
2. Experience (yrs post-PhD)	
<3 yr	5%
3–5 yr	18%
>5 yr	20%

research performance to national economic goals, which requires a dramatic expansion of Australia's capacity to produce graduates, particularly in science and technology. The success of this strategy depends on the ability to attract the brightest young minds to careers in scientific research.

Too much secrecy

SIR—I would like to add to the letter from A. Thyagaraja (*Nature* 335, 391; 1988) some observations based on my own experience of writing and refereeing papers over the past 30 years or so.

On the relatively rare occasions when I have received signed referees' reports, they have been cogently expressed and I have welcomed them, irrespective of whether they were favourable or not. Unfavourable anonymous reports are less welcome. I have often found them difficult to reply to, because they tend to be cryptic if not incomprehensible; anonymity seems to invite a casual approach.

At the risk of sounding self-righteous, I cannot recall ever writing an anonymous report; my name may have been removed by the editor in some instances, but not at my request.

If I think a paper is unsatisfactory by reason of, for example, length, confused presentation, insufficient or excessive supporting data or inadequate diagrams, I write to the authors wherever possible and send a copy to the editor. Only once did an author take exception to this. He complained to the editor, who adjudicated (in

A fundamental flaw in the government's new policy is its failure to address existing problems in the training and retention of research scientists in productive research careers. Cutbacks in research and higher education, spending over the past ten years have significantly reduced the number of continuing positions available, leading to a 'bottleneck' between non-tenured and continuing appointments. Young scientists are forced either to spend longer in short fixed-term positions or to leave research and seek employment in fields that offer better prospects for career advancement.

In July–August 1988, we conducted a survey of the attitudes of non-tenured scientists at the Australian National University (ANU) in Canberra. A total of 194 non-tenured research staff (94 per cent of all non-tenured scientists at the ANU) were surveyed, spanning the disciplines of physics (24 per cent), chemistry (21 per cent), biology (37 per cent), medicine (10 per cent), Earth sciences (4 per cent) and mathematics (4 per cent). The response was high (70 per cent). The respondents had varying amounts of postdoctoral experience, ranging from less than three years (29 per cent), 3–5 years (32 per cent) and greater than five years (39 per cent). An overwhelming majority of non-tenured scientists were pessimistic about their future prospects in Australian research, describing them as "poor"

my favour, as it happens, but at least the author had an identifiable target).

In nearly every case, authors have responded thoughtfully and courteously, adopting some of my suggestions, rejecting others, always with sound explanations for their actions. I think the result has been to improve the quality of the papers. (Sometimes I even get a mention in the acknowledgements.) Where it has been necessary to reject a paper, I have always provided a full explanation, so that the author(s) can come back to me direct.

Nobody has yet presented me with a convincing argument in favour of anonymous refereeing, and that includes practising scientists as well as journal editors. Quite simply, I do not believe there is one.

Most vulnerable to anonymous refereeing are those working in controversial or rapidly developing fields, especially if they are young and trying to make a name for themselves. Show me a scientist who favours anonymous refereeing and I will show you someone who is insecure. That is my opinion, after some 30 years' experience. What do other people think?

J. B. WRIGHT

Department of Earth Sciences,
Open University, Walton Hall,
Milton Keynes, MK7 6AA, UK

(Table 1). The survey indicated a very high level of dissatisfaction with career structures at present available to Australian researchers (Table 2). Sixty-one per cent of respondents were prepared to leave Australian research (Table 2). The Australian National University is often regarded as the 'showpiece' of the Australian university system with respect to its research effort. We believe the results of this survey reflect the situation faced by the majority of non-tenured scientists in Australian universities.

Our survey indicates one likely consequence of a continuing lack of career opportunities in Australian research. In the near term, many talented young scientists will leave the Australian system (see Table 2), either to continue research careers in other countries, or to find employment with better career prospects outside research. Another likely conse-

Table 2 Attitudes
Dissatisfaction with present career structures in Australian science

1. Overall	91%
2. Nationality	
Australian	96%
non-Australian	73%
3. Experience (yrs post-PhD)	
<3 yr	76%
3–5 yr	93%
>5 yr	98%
Prepared to leave Australian research	
1. Overall	61%
2. Nationality	
Australian	59%
non-Australian	78%

quence is that high-quality graduates will be dissuaded from pursuing careers in scientific research. In the longer term, this loss of talent will result in Australia's research and development effort failing to provide the technological base for the planned revitalization of the Australian economy.

Solutions can be found to the problems highlighted by our survey. It is vital to increase funding for research and higher education. Moreover, the first step must be to address the existing problems with career structures. More continuing appointments, based on merit, must be made to retain young scientists and to recruit prospective students to careers in research.

M. K. MORELL

G. D. PRICE

K. C. WOO

K. T. HUBIC

A. B. WELLINGTON

Research School of Biological Sciences,
Australian National University,
PO Box 475,
Canberra,
ACT 2601,
Australia

Pulsar shapes explained at last

A new classification of the shapes of radio pulses from pulsating stars is a reminder that some matters remain unresolved even after two decades of energetic research.

THE temptation to forget that nearly 21 years have passed since the discovery of pulsating stars (see *Nature* 217, 709; 1968) can only be reinforced by the simple truth that people in the field are still hesitant of being dogmatic about the mechanism of radio emission from these objects. All this is curiously in contrast with what seemed the helter-skelter deepening of understanding that followed the first discovery.

Then, as this journal knows all too well, people were talking of emission from dwarf or even neutron stars within weeks of the discovery of the first pulsar at Cambridge. Professor Thomas Gold was the first to spell out the doctrine in detail. Within a year, there was evidence from timing observations that the spin of some neutron stars could unpredictably accelerate, presumably as a consequence of the rearrangement of solid material on the surface. The idea that spinning neutron stars will usually be magnetized, with surface field strengths unimaginably greater than those available on the surface of the Earth, was also an easy step forward, accounting among other things for the mostly uniform deceleration of the spin of a pulsar. Yet authors in the field have seemingly been hard-pressed to account for the shapes of the radio pulses observed on the surface of the Earth.

Much of the difficulty is diversity: The shape of the pulse is the variation of the radio-emission intensity as a function of phase in the repetitive period of the pulsar. Historically, it fell to people at Jodrell Bank, the radioastronomy observatory at the University of Manchester, to produce the first pulse shapes. To begin with, as even now, they were surprising and even puzzling. While the pulses from some rotating neutron stars are what might be expected — simple gaussian-like excesses of radio-emission — others were double peaks concentrated in one part of the duty cycle while others were still more confused, with a small 'inter-pulse' somewhere between the main emissions. To begin with, this variety of shapes seemed to offer a means of explaining what was happening at the surface of a pulsating star, but the diversity was such that people were reduced to attempts merely at classification.

Only now, after twenty years, does the process of classification seem to make a coherent tale. Appropriately, one of those responsible is A.G. Lyne from Jodrell Bank, one of the group that discovered

the second known pulsar. With R. N. Manchester, another pulsar observer of long standing now at the CSIRO Division of Radiophysics at Epping (near Sydney) in Australia, he has now produced the most detailed classification yet of the shapes of pulsar pulses (*Mon. Not. R. astr. Soc.* 234, 477; 1988) based on analysis of the pulses from more than 200 pulsars.

The essence of the case is simple enough. A spinning neutron star produces a cone-shaped pattern of radio-emission whose axis coincides with the magnetic dipole axis of the neutron star. The magnetic and rotational axes of the star do not necessarily coincide, which means that the cone-shaped pattern of radiation is swept across the sky at the frequency of the rotation of the star; the often-quoted simile of the lighthouse-beam applies. Sharp single pulses are then observed either when the cone is extremely narrow and uniformly filled with radiation or, alternatively, when the line of sight from an observer on the surface of the Earth to the magnetic pole of the pulsating star just happens to graze the edge of the cone.

On this simple but correct picture, radio pulses with two peaks arise because the distribution of radio-emission with the revolving cone is far from uniform, and in particular is depleted along the magnetic axis relatively to larger angles with it. Then the two peaks arise as the line of sight intercepts first one direction on the revolving cone and then another, perhaps a whole diameter away.

With the passage of twenty years, there is now a wealth of detailed information with which to work. Lyne and Manchester use, in particular, two kinds of information not available in the early days — measurements of the polarization of the radio-emission (gathered as a by-product in attempts to estimate interstellar electron densities from the Faraday rotation of the plane of polarization) and the differences between the pulse profiles at different radio frequencies.

The conclusions are gratifying in that they rationalize diversity, allowing different and complicated pulse shapes to be reconciled with a simple underlying model. The notion that a cone of emission is regularly swept across the sky with the rotational frequency of the underlying neutron star is familiar. So too is the much earlier supposition that the marked separation of some double peaks suggests

that the conical pattern of radio-emission from a magnetic pole of the underlying star is often concentrated towards the outer boundaries of the cone.

Surprisingly, given that the mismatch between the rotational and magnetic axes of the star is plainly asymmetric, there is no evidence from the observations that the cone is anything but circular. And the opening angle of the cone appears to be strictly a function of the pulsar period — the more rapid the rotation, the wider the cone. Alternatively, the angle of the cone is inversely proportional to the cube root of the pulsar period.

In the pulsars with a marked concentration of radio-emission towards the outer parts of the cone, it seems that there is very little regularity of the distribution of radio-emission within the extremities. Instead, as Lyne and Manchester put it, there are patches of radiation, apparently stable enough to be measured over a great many duty cycles. And while there may be apparently chaotic variations of radio-emission within the opening angle of the revolving cone, Lyne and Manchester see no reason to suppose that the mechanisms of radio emission from the two regions are different.

In all this, perhaps the most surprising feature of the shape of pulsar pulses is that the emission usually derives from just one of the two magnetic poles. That, after all, is why pulsating stars with inter-pulse pulses are the exceptions, not the rule. In the event, it seems, the second pole is active only in the youngest pulsating stars, either those newly formed in supernova explosions or those which have been newly given a high rotation speed by the accretion of material from a companion star — now the favourite explanation of the origin of the pulsars with rotational periods measured in milliseconds.

Lyne and Manchester thus seem to have brought the classification of pulsar shapes to a satisfactory conclusion. But there is more than mere stamp collecting in the endeavour, for the relationship between the opening-angle of the emission cone and the rotational period of a pulsar is a practical guide to the chance that a particular neutron star will be detected on the surface of the Earth. For the greater the angle, the greater the chance that its cone will sweep across the Earth, and therefore the smaller the chance that observation will be biased against its detection.

John Maddox

Planetary science

Magnetic fields from impacts

Mike Fuller

ONE of the most spectacular discoveries of these early years of the age of space exploration is the ubiquitous record of cratering on ancient planetary surfaces. For those who seek to document the magnetic fields of the early Solar System, the cratering is of great interest because the impacts may have generated magnetic fields. Furthermore, if such impact-generated magnetic fields could explain observed magnetic results, planetary scientists would not have to invoke internally generated fields, dynamos and cores to explain the palaeomagnetic records of planets. Despite this interest, definitive laboratory-scale experimental work has been lacking in what is clearly a difficult area. But on page 50 of this issue¹, Crawford and Schultz report intriguing results on the generation of magnetic fields by low-angle, hypervelocity impacts in a low-field environment.

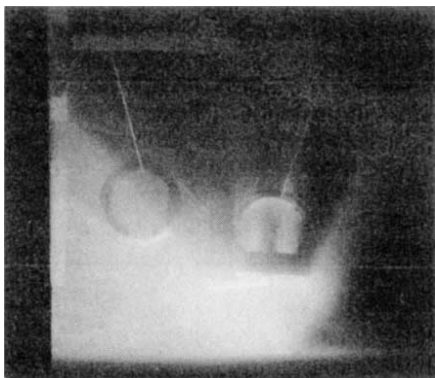
Their work was carried out using the NASA Ames vertical gun range, which fires centimetre-size metallic or plastic projectiles into an evacuated chamber at velocities of 3–7 km s⁻¹. The gun is mounted on a large A-frame so that it can fire through different ports to provide various impact angles. The authors carried out the experiments in a reduced field of about 500 nanotesla (nT) and at various higher field strengths. For comparison, the present field of the solar wind is a few nanoteslas. A powder target and a low angle of impact were used to enhance the production of a partially ionized cloud (see figure).

The authors monitored the magnetic fields with three coil arrays. The impacts generate vertical fields of up to 100 nT beneath the target in each experiment, regardless of the sign or magnitude of the ambient field. A vertical coil monitored toroidal fields of 100 and 1,000 nT generated above the target. These fields are also independent of the ambient field. The vertical coils down-range from the target measured compressed-ambient-field signatures which reversed in sign with the ambient field. The authors conclude that the spontaneous fields in the target area could be produced in a field-free environment by hypervelocity impacts, a conclusion which could be confirmed by more detailed field monitoring with devices such as miniaturized Hall probe detectors. The fields arise from currents which flow in the plasma in response to strong temperature and pressure gradients. This field generation is to be distinguished from field amplification by plasma compression (which I discussed in a previous News and Views article²).

These experiments join a substantial,

but not yet definitive body of work which seeks to establish whether impacts generate magnetic fields^{3,4}; whether such fields magnetize target materials; and whether this is important on planetary surfaces.

Experiments clearly show that impacts can magnetize materials when performed in ambient fields of the order of microteslas. Shock magnetization is the dominant mechanism of magnetization in material not heated substantially and



Luminous vapour cloud produced by a 15° impact at 4.76 km s⁻¹ of a 0.64 cm aluminium projectile on a powdered dolomite target. Magnetic search coils buried beneath and suspended above the target measured the magnetic signals generated during the event.

thermoremanent magnetization is acquired by material whose residual temperature is sufficiently high. In addition, new magnetic phases generated will acquire magnetization. It is less clear whether the fields discussed by Crawford and Schultz would be recorded. If the field were to coincide with the shock wave passing through the material, it could produce shock remanent magnetization. Indeed, the proposed vertical field may thus explain observations of shock magnetization made 15 years ago in our laboratory⁵.

It is also possible, as Crawford and Schultz suggest¹, that microscopic fragments heated by the impact might cool in the impact generated fields. But it seems unlikely that fields extant only for a few milliseconds could magnetize any substantial mass by this mechanism because the cooling would be insufficient.

As for the planetary effects of impact magnetization, impact craters on Earth have associated magnetic anomalies. Thus, in the geomagnetic field of tens of microteslas, crustal magnetization does take place. As a corollary, the absence of magnetic anomalies at a crater indicates strongly that the impact did not take place in the presence of such fields. There are abundant craters on the Moon. But studies of the magnetic fields associated

with them give little comfort to those who favour impact-generated magnetization. Lin has convincingly shown⁶ that the cratered regions are not in general more magnetic than the non-cratered. Indeed, the only basins and craters which have associated magnetic anomalies were formed when the Imbrium and other great basins were formed⁷, suggesting that only at this time (3.6–3.9 × 10⁹ years ago) did the Moon's surface experience a strong magnetic field of around a microtesla. It seems that the magnetization from fields generated by impacts could account for the magnetization of some lunar samples, but is unlikely to be a significant source of lunar magnetic anomalies.

Crawford and Schultz's suggestion that the magnetized glass on lunar sample 70019 records thermoremanent magnetization acquired in an impact-generated field is plausible, but not compelling. The sample was collected to see whether glass generated in a very young crater carries magnetization. If the scaling for the 3-m crater in which the glass was found is correct, the sample must have cooled in milliseconds to match the data. This has previously led Sugiura and colleagues to suggest that the glass might have come from a nearby larger crater⁸. Moreover, although the actual intensity determinations represent a noble effort to overcome the difficulties encountered in working with the lunar samples, they do not satisfy standard quality control tests of the method. Such questions should be studied in a long overdue re-examination of lunar samples with modern techniques.

The intriguing experimental results of Crawford and Schultz once again raise the issue of the importance of impacts as a source field for the magnetization of ancient planetary surfaces. Compression and amplification of ambient fields has been used to explain planetary surface magnetization, as a result of impacts of solid bodies^{9,10} or comets¹¹. This is likely to be an important mechanism. These new experimental results from Crawford and Schultz provide further evidence that the generation of magnetic fields by impacts could provide a mechanism for magnetization in very low-field environments. It remains to be seen whether the strong temperature and pressure gradients that drive the field generation can be scaled to large impacts on a planetary surface. □

1. Crawford, D.A. & Schultz, P.H. *Nature* **336**, 50–52 (1988).
2. Fuller, M. *Nature* **329**, 674–675 (1987).
3. Pert, G.J. *J. Plasma Phys.* **18**, 227–241 (1977).
4. Martelli, G. & Newton, G. *Nature* **269**, 478–482 (1977).
5. Cisowski, S.M. *et al. Proc. 7th Lunar Sci. Conf.* 3299–3320 (1976).
6. Lin, R.P. *Phys. Earth Plan. Int.* **20**, 271–280 (1979).
7. Russell, C.T. *et al. Proc. 8th Lunar Sci. Conf.* 831 (1977).
8. Sugiura, M. *et al. Proc. 10th Lunar Sci. Conf.* 2189 (1979).
9. Hide, R. *The Moon* **4**, 39 (1972).
10. Hood, L.L. *Geophys. Res. Lett.* **14**, 844–847 (1987).
11. Gold, T. & Soter, S. *Planet. Space Sci.* **24**, 45–54 (1976).

Mike Fuller is at the Department of Geological Sciences, University of California, Santa Barbara, California 93106, USA.

D. A. Crawford/NASA Ames

Evolution

No barriers to speciation

N.H. Barton, J.S. Jones and J. Mallet

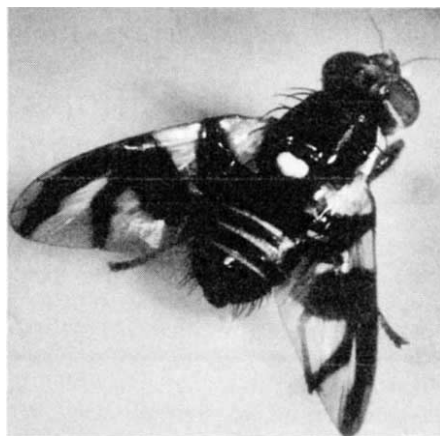
THE *Origin of Species* is mainly not about the origin of species as, in Darwin's view, speciation was simply an extension of morphological evolution within existing lineages. Now that species are seen as genetically isolated units, their origin is harder to understand as it must involve the appearance of individuals unfit because they are less able to exchange genes with others. Most evolutionary biologists believe that new species can evolve only if there is a geographical barrier to the exchange of genes between ancestor and descendant; that is, in allopatry. The appearance of a new species in the same place as its predecessor (in sympatry) or in an adjacent population (parapatry) is usually dismissed on the grounds that gene flow and selection against hybrids will swamp any tendency to reproductive isolation. In contrast, population genetics theory shows that under certain conditions of natural selection and population structure, mating barriers can evolve despite gene flow¹.

The intellectual allopatry of theoreticians and naturalists has allowed these views to develop in isolation, but several studies of populations in nature can now plausibly be interpreted in terms of sympatric differentiation. Feder *et al.*², McPherson *et al.*³ and Smith⁴ in three papers in this issue investigate the evolution of host races in the apple maggot fly *Rhagoletis pomonella* (see figure). This fly is native to hawthorn in North America, but in the nineteenth century spread to introduced apple trees, and has now begun to infest cherries, roses and pears. It provides a unique chance to study the history and geography of evolution, and its spread to a new and sympatric host does seem to be reflected in behavioural, ecological and genetic changes which might be a precursor of speciation.

Apple and hawthorn flies differ in host preference. The offspring of wild females collected from apples are much more likely to select these fruits as an egg-laying site than are the descendants of hawthorn flies⁵. As apple males also prefer to rest (and hence presumably to mate) on their native fruit, there is the opportunity for reproductive isolation between the races. This behavioural barrier is strengthened by a difference in developmental time. Apple flies take about 40 days to mature in the laboratory, whereas those from hawthorn (whose fruit ripens later in the season) do not become adult for a further two to three weeks⁴. Crosses between them show that the timing differences are inherited, but give no evidence of

reproductive mating barriers⁵.

Such differences in habitat choice can be an effective impediment to gene flow, and many good species are separated by nothing more than a habitat or temporal preference which ensures that they do not meet to exchange genes. In animals which mate on or in their preferred host, positive assortative mating is an inevitable by-product of habitat choice. The process has been simulated in experiments on *Drosophila*: flies selecting different light, gravity and chemical cues in a maze rapidly diverge into genetically isolated populations mating only with others sharing their own habitat preference⁶. Behavioural and temporal barriers in *Rhagoletis* are reflected in genetic divergence, as in both Illinois and Michigan there are marked differences in the frequency of enzyme alleles between the apple and hawthorn



Rhagoletis pomonella—speciation in progress? (Courtesy of J. L. Feder *et al.*)

racess (with, oddly enough, some of the patterns of association changing from place to place)^{2,3}. As the flies can move much further than the distance between the apple and hawthorn trees, there is a case to be made here for sympatric divergence of host races.

Within each race, there are strong associations (linkage disequilibria) between the divergent enzyme alleles. Population-genetics theory allows us to use these to calculate the probable extent of gene exchange between the two races⁷. Feder *et al.* show², for example, alleles at the *Dia-2* and *Aat-2* loci (which have a recombination distance $r = 3.2$ per cent) have an association in the Michigan population which is about half of the maximum possible (standardized disequilibrium $R = 0.53$). Because recombination alone would eliminate these associations in much less than the hundred or so generations during which the races have been

diverging (with an expected disequilibrium of $(1 - 0.032)^{100} = 0.039$; much less than that observed) something must be maintaining them.

The most likely candidate is gene flow. The continual introduction of chromosomes from one race into the other will maintain allele associations in the face of recombination ($R \approx m\Delta p_1\Delta p_2 / r\sqrt{p_1q_1p_2q_2}$) and these associations give an estimate of gene flow, m , of 20 per cent per generation. This agrees with the measures of temporal overlap in the laboratory as the variation in developmental time between the two races⁴ would allow up to 30 per cent cross-mating. Such extensive gene flow requires correspondingly high levels of selection to maintain genetic differences between races. But other isolating mechanisms have apparently not evolved to reinforce the initial divergence. Once again, reinforcement of isolating mechanisms in response to selection against hybrids seems to be a less important phase of speciation than was once thought.

How widespread is diversification through local host or habitat specialization, and what conditions are needed to allow new species to evolve in this way? There are certainly plenty of instances of florid speciation in animals specializing on particular hosts⁸. Plant-eating insects show extreme radiation with, for example, 300 species of leaf miners in Britain. In fig wasps, whose reproduction is closely tied to the plant in which they hatch, each of the 750 species can grow only on its own species of fig⁹. In animals, too, there may be a close association between host and parasite. Sometimes these associations are the result of a history of parallel speciation in host and parasite; but in other cases (such as the closely related species of acanthostomine worms found in many fish-eating animals), parasite speciation reflects an invasion of a taxonomically distinct host analogous to that found in *Rhagoletis*¹⁰. There may be rapid evolution of the parasite within its new host: human immunodeficiency viruses, for example, have diverged considerably from their simian ancestors since the host shift some forty years ago¹¹.

If there is to be speciation by host shift, there must be evolution of genes both for host preference (and hence the opportunity for assortative mating) and for performance on the new host¹². The search for correlated changes of this kind in plant-eating insects has had mixed success. The fall cankerworm (*Alsophila pometaria*) infests various plants, but sympatric clones from different hosts show few associations between performance and preference¹³. In contrast a host-plant shift in the butterfly *Euphydryas editha* gives evidence of covariance in choice and exploitation. Some individuals in a Nevada population have incorporated

the introduced European weed *Plantago lanceolata* into their diet, whereas others avoid it¹⁴. Preference for the new food plant has a high heritability, and caterpillars develop more successfully on the host preferred by their mothers¹⁵. Although there is no evidence for reproductive isolation between butterflies with different preferences, there is at least the opportunity for the sympatric evolution of host races.

Evolutionary biology is often an attempt to reconstruct history: even for the recent past this is always difficult. In *Rhagoletis*, for example, it is hard to be certain that the apple race is not an existing sibling species which became common only after it invaded apples. Nevertheless, the new acceptability of the view that speciation does not always demand a period of geographical isolation may help

to explain the origins of the tens of millions of species estimated¹⁶ to be alive today. Perhaps the patron saint of evolution is, after all, sympatric. □

1. Kondrashev, A.S. *Theor. Pop. Biol.* **29**, 1 (1986).
2. Feder, J.L., Chilcote, C.A. & Bush, G.L. *Nature* **336**, 61–64 (1988).
3. McPherson, B.A. *et al. Nature* **336**, 64–66 (1988).
4. Smith, D.C. *Nature* **336**, 66–67 (1988).
5. Prokopy, R.J., Diehl, S.R. & Cooley, S.S. *Oecologia* **76**, 138 (1988).
6. Rice, W.R. & Salt, G.W. *Am. Nat.* **131**, 911 (1988).
7. Barton, N.H. *Evolution* **36**, 863 (1982).
8. Butlin, R. *Trends Res. ecol. Evol.* **2**, 310 (1987).
9. Kjellberg, F. *et al. Evolution* **41**, 693 (1987).
10. Mitter, C. & Brooks, D.R. in *Coevolution* (eds Futuyma, D.J. & Slatkin, M.) 65 (Sinauer, Cambridge, Massachusetts, 1983).
11. Smith, T.F. *et al. Nature* **333**, 573 (1988).
12. Felsenstein, J. *Evolution* **35**, 124 (1981).
13. Futuyma, D.J. & Philippi, T.E. *Evolution* **41**, 269 (1987).
14. Thomas, C.D. *et al. Evolution* **41**, 892 (1987).
15. Singer, M.C. *et al. Evolution* **42**, 977 (1988).
16. May, R.M. *Science* **241**, 1441 (1988).

N. H. Barton, J. S. Jones and J. Mallet are in the Department of Genetics and Biometry, University College London, London NW1 2HE, UK.

Microtubule motors

Cytoplasmic dynein graduates

Howard Stebbings

A QUARTER of a century after dynein was first isolated from 'arms' on ciliary axonemal microtubules¹, and almost as long since it was suggested to exist in multiple forms², a recent meeting* has been the first to devote a major part of its programme to the topic of cytoplasmic dynein. The intention was to explore the force-producing molecules dynein and kinesin by bringing together those who work on ciliary and flagellar function and those who work on cytoplasmic translocation. It transpires that these fields have been segregated artificially.

It has taken so long for cytoplasmic dynein to gain recognition largely because the initial experiments were carried out with sea urchin eggs^{3,4}. This created the uncertainty that the egg dynein could be a form of precursor of the dynein in the axonemes of the highly ciliated embryos of these organisms. The recent finding that a microtubule-associated protein (MAP 1C) of mammalian brain shares many of the characteristics of dynein^{5,6}, together with the identification of a dynein-like protein in an organism devoid of cilia and flagella at any stage of its life cycle⁷, has resurrected and renewed interest and excitement in the subject. The meeting featured cytoplasmic dyneins, not only from sea urchins and vertebrate brain, but also from other tissues such as liver and testis as well as from various cultured cell types.

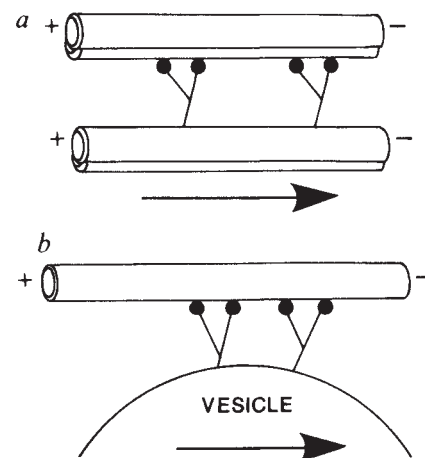
As one might imagine, investigations of cytoplasmic dyneins have benefited greatly from what is known of axonemal dyneins, and methods and techniques

which were fruitful in studying axonemal dyneins are being applied to their cytoplasmic counterparts. It is now likely that research into the different forms of dynein will progress hand in hand.

Structural, immunological, biochemical and functional comparisons between dyneins from different sources, and between dyneins and the other characterized microtubule-associated mechanized enzyme, kinesin, featured conspicuously at the meeting. The common properties are ATP-sensitive binding to microtubules, microtubule-stimulated ATPase activity, and the promotion of unidirectional movement or components along microtubules or movement of microtubules relative to each other or to glass substrates.

Dynein, a complex molecule, is comprised of several subunits. It has been known for some time that differences exist even between the relatively few axonemal dyneins which have been studied in detail. Outer-arm dynein from cilia and flagella from the protozoans *Tetrahymena* and *Chlamydomonas* is composed of three subunits, whereas in sea urchin axonemes it is composed of two. It now turns out that trout sperm and chicken tracheal axonemal outer-arm dynein also have two subunits, suggesting that the protozoans may be an exception. And cytoplasmic dynein from brain of chicken shows a similar morphology (E. R. Steuer, Washington Univ., St Louis) with two globular heads joined by stems. The only discernible difference in the case of cytoplasmic dynein is that both heads appear spherical, whereas in ciliary dynein one is spherical and one pear-shaped.

Each dynein subunit contains a single heavy chain and one or more intermediate and/or light chains. These chains can differ in molecular mass even between cilia and flagella of the same organism, certainly between axonemal and cytoplasmic dyneins and probably between the cytoplasmic dyneins. These differences seem to reflect the existence of multiple dynein isoforms. The differences between the intermediate and light chains may be involved in the specificity of dynein binding to cellular organelles. Immunological comparisons between the dynein heavy



a, Axonemal dynein bringing about a relative sliding of microtubule doublets in cilia and flagella, and b, how cytoplasmic dynein might interact with a single microtubule to cause retrograde movement of the organelle to which it is bound.

chains from different sources have produced conflicting results. In one study (D. J. Asai, Purdue Univ.), antibodies raised against heavy chains of sea-urchin dyneins from sperm flagella, unfertilized eggs and embryonic cilia reveal that all three are homologous, but that each can be distinguished from the other. Ironically, this study supports the original suspicion* that at least one egg dynein is a ciliary dynein precursor.

Dynein subunits have an ATPase activity associated with the heavy chain, but the enzymology of axonemal and cytoplasmic dyneins is somewhat different⁹. The former hydrolyses only ATP whereas the latter can hydrolyse other nucleotides, even though these do not produce movement. Flagellar and cytoplasmic dynein ATPases are strongly inhibited by vanadate, and of particular interest are details of the mechanism of photocleavage of dynein heavy chains by irradiation with ultraviolet light in the presence of vanadate, with and without ATP. The phenomenon, which is regarded as diagnostic of dynein, has been used successfully to probe the structure of various dyneins, and in particular to localize the ATP-binding and the microtubule-binding sites along the heavy chains (S. M. King, Worcester Foundation for Experimental Biology, Massachusetts).

* Force Production and Microtubule-Coupled Cell Movement, satellite meeting of the International Congress of Cell Biology, Stowe, Vermont, 20–25 August 1988.

Most important, the technique has been crucial in the identification of cytoplasmic dyneins from minor sources and for comparing them with ciliary and flagellar dyneins, and it is also being used as a functional probe applied to cytosol fractions and extracted cell models.

The properties of axonemal dyneins have been investigated using *in vitro* assays developed to study microtubule translocators. All dyneins produce movement towards the minus ends of microtubules (retrograde), which is opposite to that brought about by kinesin (see the previous News and Views article by Ian Gibbons¹⁰). The methods have gone beyond simply allowing the demonstration of the ability of a putative motor to produce movement, and thence the determination of its polarity. They are being used to ascertain whether and which individual subunits of dynein are capable of producing movement, and to gain information on the molecular mechanisms of movement generated by different motors, by accurately analysing precise differences in movement driven by microtubule-based motor enzymes. For example, it has been shown that when three-headed outer-arm dynein is proteolytically digested into two- and one-headed particles, the former can produce movement whereas the latter cannot; and with two-headed dynein one subunit, the β -heavy chain, is sufficient to bring about movement, indicating that a single domain is all that is required (R. D. Vale, University of California). Using similar techniques, bead movements driven by dynein and kinesin have been seen to be distinctly different, suggesting that the two motors interact with the microtubule surface differently as a result of different molecular motility mechanisms (J. Gelles, Washington Univ., St Louis).

Dynein is known to couple a cycle of ATP binding and hydrolysis with a mechanical binding and release cycle to produce a relative sliding of microtubule doublets in axonemes. Although axonemal dynein motility seen *in vitro* can be equated readily with the *in vivo* function, the same is not so easy with cytoplasmic dyneins. The discovery of dynein in neuronal tissue has led to the suggestion that it is in fact the retrograde motor for axoplasmic transport¹¹ (see figure). Its role in translocation *in vivo* has been supported by its identification in a range of systems such as fish chromatophores and insect ovaries in which translocation is greatly pronounced. It has even been suggested that dynein or a similar molecule is involved in mitosis.

Are there other microtubule motors? Already, R. B. Vallee and co-workers (Worcester Foundation) believe that they have found another mechanochemical enzyme, distinct from both dynein and kinesin, in calf brain cytosol. This protein

has a relative molecular mass of 100,000, shows ATP-sensitive binding to microtubules and may have ATPase activity. It bundles microtubules via crossbridges, and is relatively insensitive to vanadate. It may therefore be the first of many more such molecules.

Debate continues as to why dynein is so complex and, as there is clearly some redundancy, why it is so well conserved. Are axonemal and cytoplasmic dyneins products of the same genes? Now that cytoplasmic dynein has become respectable it is unlikely to be a further 25 years before these topics are discussed. □

1. Gibbons, I. R. *Proc. natn. Acad. Sci. U.S.A.* **50**, 1002–1010 (1963).
2. Weisenberg, R. & Taylor E. W. *Expt Cell Res.* **53**, 372–384 (1968).
3. Pratt, M. M. *Dev Biol.* **74**, 364–378 (1980).
4. Hisanaka, S. & Sakai, H. *Dev Growth Differ.* **22**, 373–384 (1980).
5. Paschal, B. M., Shpetner, H. S. & Vallee, R. B. *J. Cell Biol.* **105**, 1273–1282 (1987).
6. Vallee, R. B., Wall, J. S., Paschal, B. M. & Shpetner, H. S. *Nature* **332**, 561–563 (1988).
7. Lye, R. J. *et al. Cell* **51**, 309–318 (1987).
8. Porter, M. E. *et al. J. biol. Chem.* **263**, 6759–6771 (1988).
9. Shpetner, H. S., Paschal, B. M. & Vallee, R. B. *J. Cell Biol.* **107**, 1001–1009 (1987).
10. Gibbons, I. R. *Nature* **330**, 600 (1987).
11. Paschal, B. M. & Vallee, R. B. *Nature* **330**, 181–183 (1987).

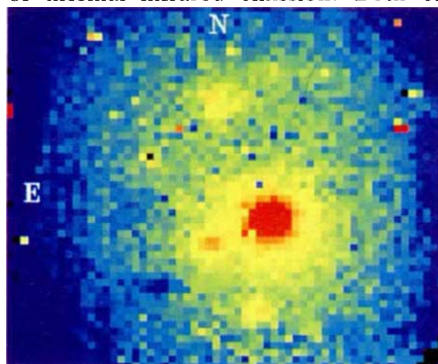
Howard Stebbings is in the Department of Biological Sciences, University of Exeter, Perry Road, Exeter EX4 4QG, UK.

Astrophysics

A clearer view of active galaxies

Martin Ward

THERE are two important areas of active-galaxy research in which recent progress has been particularly rapid. The first concerns the ionizing continuum emission from the nucleus — whether we are seeing this radiation directly. The second deals with the controversy over the importance of thermal infrared emission. Both of



The central 8,000 light years of the active galaxy NGC4151, obtained using the infrared camera on the UKIRT, through a filter sensitive to the hydrogen emission line, Brackett- γ . The bright nucleus (red) is surrounded by a patchy distribution of dusty obscured regions of ionized hydrogen (H II, yellow), which could contribute significantly to the thermal infrared continuum. (Courtesy UKIRT IRCAM team.)

these issues were scrutinized closely at a recent symposium*.

A fundamental question about active galaxies is: are we seeing all of the radiation from the nucleus, or merely some fraction that depends on our specific viewpoint? In a recent News and Views article (*Nature* **334**, 569; 1988) I discussed work by Alighieri and co-workers that strongly indicates collimation of the ionizing radiation from a radio galaxy. At the meeting diverse evidence was presented that the continuum emission from many active galaxies is similarly anisotropic. Polarization studies (J. Miller, Lick Observatory)

show that the reflection model for Seyfert-2-type active galaxies, which I discussed in my article, applies generally rather than being a rare quirk. According to this model, we see only scattered radiation, representing a small fraction of that emitted from a nucleus which is hidden from direct view.

Other evidence comes from pictures of nuclei taken through filters sensitive to different spectral emission lines. Combined, these images reveal areas of high atomic excitation. The line-ratio maps define the galactic regions that are irradiated directly by the nucleus (R. Pogge, University of Texas). For one galaxy, NGC1068, this technique shows that there is a cone of highly excited gas, with its apex pointing towards the nucleus. Independent evidence is obtained by analysing the motion (wavelength shifts) of the line-emitting gas distributed around the nucleus. The results for NGC4151 indicate (H. Schulz, Ruhr University) a hole in the distribution of obscuring material around the nucleus of this well known Seyfert galaxy, through which ionizing photons can escape, producing a plume of optically excited gas extended in one direction. The image in the infrared is quite different (see figure) because we can see the obscured ionized gas.

The emission lines from the dense, high-velocity gas close to the nucleus can change in strength in less than a week, even in luminous objects. Estimates of the size of the region can be made from the distance light travels over such a period (M. Gaskell, University of Michigan). It has been noted that this size often conflicts with the larger sizes inferred from photoionization models (B. Peterson, Ohio State University). But it was suggested that this problem may be illusory and can be removed by using more appropriate mathematical methods to deduce the true size of emission line regions from variability

*IAU Symposium no. 134 *Active Galactic Nuclei*, University of California, Santa Cruz, 15–19 August 1988.

data (R.A. Edelson, University of Colorado). A huge collaboration has been initiated to monitor emission lines in selected active nuclei using the International Ultraviolet Explorer (IUE) satellite from next year. The hope is that these concerted observations with only small time gaps will permit much more realistic size estimates of the line-emitting region.

It is also possible that the size problem arises from the anisotropic continuum emission (M. Penston, Royal Greenwich Observatory). Because in some objects the collimation points nearly towards us, perhaps the emission line region does the same. For this geometry the region seems foreshortened from our viewpoint and the size deduced from line variations will be smaller (depending on the angle) than the true dimension of the region. Collimated emission is familiar to radio astronomers, but its significance at other frequencies is now beginning to be appreciated.

The other area of progress discussed concerns the origin of the infrared to millimetre continuum radiation of Seyfert galaxies and quasars. Early studies of quasars were driven primarily by radio-frequency observations, for which synchrotron emission is often the dominant process. Because of this bias, it is perhaps not surprising that non-thermal models have been widely used to explain the continuum emission at other frequencies, particularly the infrared. Although in strong radio sources the infrared continuum is probably dominated by non-thermal emission, the situation for the 90 per cent of active galaxies that are radio quiet is still controversial. Just as radio observations indicated the dominance of non-thermal processes, so recent IRAS (Infrared Astronomical Satellite) observations in the far infrared (B. Soifer, Californian Institute of Technology) and others at millimetre wavelengths (A. Lawrence, Queen Mary College; E. Krugel, Max Planck Institute for Radioastronomy) reveal the importance of thermal emission. Many of the millimetre-wavelength observations that bridge the gap between the radio and infrared detections come from new facilities such as the James Clerk Maxwell telescope in Hawaii and the IRAM (Institut d'Radioastronomie Millimétrique) telescope in Spain. Millimetre observations are critical because they define the change in shape of the continuum flux distribution between infrared and radio wavelengths. Generally, models with non-thermal components predict a less pronounced decrease of flux with increasing wavelength than would be observed if the dominant component was thermal.

In many, if not all, of the radio-quiet nuclei, the infrared to millimetre continuum is consistent with thermal emission which could arise from dust grains heated by radiation from the nucleus or from star-

forming regions surrounding the nucleus. Infrared images could confirm the latter possibility. Using the new near-infrared array detectors, we can directly see some dust-obscured ionized-hydrogen (H II) regions surrounding the nucleus of NGC4151 (see figure). It is quite probable that there is a mixture of thermal and non-thermal components in the same nucleus and the observations do not preclude an underlying non-thermal component. But

to distinguish this from a much stronger thermal source at millimetre and far-infrared frequencies will require observations with high spatial resolution because the thermal source will be extended and the non-thermal source, compact. Such resolution is not yet attainable at the crucial far-infrared wavelengths. □

Martin Ward is at the Institute of Astronomy, Cambridge University, Madingley Road, Cambridge CB3 0HA, UK.

Cell physiology

Answer in a flash

David Ogden

MANY processes in cell physiology will be understood only when the rates and equilibria of the reaction steps involved are known. One way to elucidate these rates is quickly to change the concentration of a ligand or substrate and to observe the rate of re-equilibration of the system. But rapid and precise concentration changes are difficult to achieve at the cell surface or in the cytosol because of slow access rates governed by diffusion, binding and inactivation in unstirred spaces. A technique designed to achieve fast concentration changes for a wide range of

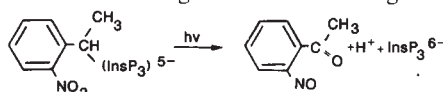


Fig. 1 Photolysis of 2-nitrophenylethyl phosphate esters. The initial photoreaction generates a proton and an acinitrile intermediate, which decays over a few ms to release free phosphate compound and 2-nitrosoacetophenone. The latter may be toxic at high concentration and can be neutralized with reducing thiols.

ligands was the subject of a recent meeting*. The cell or tissue is pre-equilibrated with an inactive, photolabile 'caged' precursor and the ligand released at its site of action by a brief light flash from a laser or flashlamp (see ref. 1 for review).

Several factors have contributed to current interest in this technique: a wide range of well-characterized caged neurotransmitters and intracellular probes have been synthesized and many are available commercially; inexpensive flashlamps suitable for photolysis in biological systems have been developed², or are in development (H. A. Lester, California Institute of Technology); and methods for introducing membrane-impermeant caged molecules into cells, such as permeabilization techniques, micro-injection or perfusion with glass micropipettes. Further advances will probably result from design and synthesis of new cages to increase the rate of photolysis; extending the range of caged ligands to include peptides, eicosanoids and caged antagonists; and

improving techniques such as photoactivation of fluorescent labels.

Suitable caged molecules should have no intrinsic biological activity, neither mimicking the active agent nor interfering with associated enzyme systems. The rate and efficiency of photolysis should be high and occur at wavelengths where intense illumination does not damage cells. The by-products of photolysis should not be toxic. The 2-nitrophenylethyl esters of phosphates (Fig. 1), including ATP, ADP, GTP γ S, cAMP, cGMP, and inositol 1,4,5-trisphosphate (InsP₃) have several desirable properties. Recent improvements in synthesis of these molecules³ and studies of their mechanism of action were described by D.R. Trentham (NIMR, London; see also ref. 4 for a review). Photolysis occurs at suitable wavelengths, 300–360 nm, rates of photolysis are fast (half time 2–3 ms) and quantum efficiency is high. Caged InsP₃, with three phosphates available for substitution, also requires separation of isomers caged on different phosphates and characterization of their biological activities⁵.

A different approach has been necessary in developing 'caged' inorganic ions such as calcium. Photolabile chelators which change from high- to low-affinity forms on photolysis release free ions following a light flash. As well as undergoing fast and efficient photolysis, these cages should be selective for the ion of interest and show a large affinity decrease on photolysis to maximize release. A new caged calcium chelator, DM-Nitrophen, based on methoxynitrophenyl derivatives of EDTA, is cleaved into two diacetic acid derivatives on photolysis, thus producing a marked reduction of affinity (J. Kaplan, University of Pennsylvania). The high initial affinity for calcium (dissociation constant 5 nM) allows a pulse of free calcium ions to be produced from normal resting cell concentrations, with recovery determined by intracellular calcium buffering. The speed of calcium release from DM-Nitrophen is illustrated by the rate of block by external calcium ions of the

* Caged compounds and their applications in biology. National Institute for Medical Research, London, 21 July 1988.

sodium current through calcium channels, which occurs within 1 ms of the light flash⁶. But DM-Nitrophen has a relatively high affinity for magnesium ions, and there are concomitant changes of magnesium concentration in the physiological range.

A series of cages based on the chelator BAPTA, which has a high selectivity for calcium over magnesium, have been developed by Tsien and co-workers⁷. These compounds, Nitr5 and Nitr7, have affinities for calcium in the physiological range (dissociation constants 150 and 50 nM) and a smaller affinity change on photolysis than DM-Nitrophen. They therefore

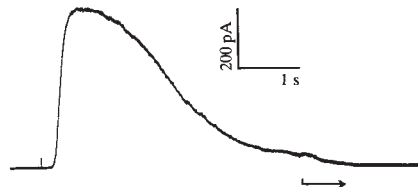


Fig. 2 Activation of Ca-dependent K conductance (outward current) following photolytic release of 1 μ M InsP_3 in a voltage-clamped guinea-pig hepatocyte. Membrane potential 0 mV. Caged InsP_3 perfused intracellularly by whole-cell patch clamp method. Photolysis produced by 1 ms pulse of 300–350 nm light delivered from a flashlamp at the time indicated by the initial electrical artefact. (T. Capiod, D.O., D.R. Trentham & J.W. Walker.)

produce changes of calcium buffering from low to high concentrations in the physiological range after photolysis, rather than a pulse of free calcium ions. Nitr5 has been used with the whole-cell patch-clamp method to study activation of calcium-dependent potassium conductances in sympathetic neurons⁸ and human thyrocytes, and an adrenaline-like enhancement of calcium current in frog heart cells by intracellularly released calcium ions (A. Gurney, St Thomas' Hospital, London).

Flash photolysis has already made a significant contribution to understanding the role of ATP hydrolysis in muscle contraction and the regulation of this contraction by calcium ions. Photolysis of caged ATP (or other phosphates) inside permeabilized muscle fibres enables identification of mechanical or structural changes with particular steps in the ATP hydrolysis mechanism. Rapid detachment of cross-bridges is produced in rigor muscle by a non-hydrolytic action of released ATP. Crossbridges then re-attach and the subsequent force generation following ATP hydrolysis is associated with the release of inorganic phosphate from the myosin ATPase (Y. Goldman, University of Pennsylvania, see ref. 9). Structural changes accompanying the mechanical events resulting from caged ATP or caged calcium photolysis have been demonstrated using X-ray diffraction (R. Goody, Max-Planck Institute, Heidelberg) or by techniques such as fluorescence polarization or birefringence.

Second-messenger regulation of cell function can be studied with caged forms of InsP_3 , cAMP, GTP or calcium ions. Release of 1 μ M or less of InsP_3 in permeabilized pulmonary artery produces contraction⁹ (A.V. Somlyo, University of Pennsylvania), and in whole-cell patch-clamped hepatocytes produces an increase of calcium-dependent potassium conductance (Fig. 2), both responses occurring after a delay of about 200 ms. Activation of α -adrenoceptors in the membrane of both cell types results in the same contractile response or conductance increase but with considerably longer delays, between 2 and 20 s. Similarly, the β -adrenoceptor-mediated increase of calcium current in frog cardiac muscle has a delay of several seconds¹⁰, whereas the increase of current produced by photolysis of intracellular caged cAMP occurs in 150 ms (J. Nargeot, Montpellier).

These results suggest slow steps in the coupling of hormone receptors to the enzymes responsible for generation of the second messenger, phospholipase C or adenylcyclase. In contrast, experiments to investigate the role of G-proteins in the regulation of neuronal calcium conductance (A. Dolphin, St George's Hospital, London) show that photolytic release of the GTP analogues GTP γ S or GMP-PNP suppresses calcium current with a slow time course (1–2 min) whereas neurotransmitter action is relatively fast, suggesting facilitation of GDP–GTP exchange by neurotransmitter action¹¹.

Flash photolysis of caged neurotransmitter analogues or ion-channel blockers in the extracellular phase (see ref. 1) can overcome slow diffusional access to the cell surface, particularly acute problems in the study of synaptic pharmacology in the central nervous system. Carboxynitrophenyl derivatives of centrally acting amino-acid or amine neurotransmitters have photochemical properties similar to the caged phosphates (G. Hess, Cornell University). These compounds, along with improvements in single-cell and tissue-slice techniques, will greatly facilitate studies of mechanism in neurotransmitter physiology and pharmacology in the central nervous system, and should lead to the development of new therapeutic agents. □

1. Gurney, A.M. & Lester, H.A. *Physiol. Rev.* **67**, 583–617 (1987).
2. Rapp, G. & Guth, K. *Pflügers Archiv* **411**, 200–203 (1988).
3. Walker, J.W., Reid, G.P., McCray, J.A. & Trentham, D.R., *et al. J. Am. chem. Soc.* **110**, 7170–7177 (1988).
4. Nerbonne, J.M. in *Optical Methods in Cell Physiology* (eds de Weer, P. & Salzberg, B.) 417–455 (Wiley, 1986).
5. Walker, J.W. *et al. Nature* **327**, 249–252 (1987).
6. Morad, M. *et al. Science* **241**, 842–844 (1988).
7. Adams, S.R. *et al. J. Am. chem. Soc.* **110**, 3212–3220 (1988).
8. Gurney, A.M., Tsien, R.Y. & Lester, H.A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3496–3500 (1987).
9. Goldman, Y. *A. Rev. Physiol.* **49**, 637–654 (1987).
10. Niedergerke, R. & Page, S.G. *Proc. R. Soc. B* **197**, 333–362 (1977).
11. Dolphin, A.C. *et al. Pflügers Archiv* **411**, 628–636 (1988).

David Ogden is in the Department of Pharmacology at King's College London, Strand, London WC2R 2LS, UK.

C.V. Raman on acoustics

C. V. RAMAN, born on 7 November 100 years ago, worked at a time when many physicists seemed able to turn their hands usefully to any branch of their science. Besides his celebrated contributions to optics, in particular his discovery of the Raman effect 60 years ago, he also published extensively on acoustics. His letters to *Nature* on the subject from 1909 to 1926 are still referred to. His clarity of expression was such that non-scientists can also appreciate his ideas. I have selected extracts, to be reproduced in the next four issues, from Raman's letters which illustrate the breadth of his interest and the originality of his thought.

For this week and next, the extracts concern the 'wolf' note of stringed instruments.

J.M. Bowsher, University of Surrey.

It has long been known that on all musical instruments of the violin family there is a particular note which is difficult to excite in a satisfactory manner, and that when this "wolf-note," as it is called, is sounded, the whole body of the instrument vibrates in an unusual degree. It seems that the difficulty of eliciting a smooth note of this particular pitch is due in some way to the sympathetic resonance of the instrument. In a recent paper G.W. White has published some experimental work confirming this view. The most striking effect noticed is the cyclical variation in the intensity of the tone obtained when the instrument is forced to speak at this point. White suggests as an explanation of these fluctuations of intensity that they are due to the beats which accompany the forced vibration imposed on the resonator. The correctness of this suggestion seems open to serious criticism. For the beats which are produced when a periodic force acts on a vibrator are essentially *transitory* in character, whereas in the present case the fluctuations in intensity are *persistent*.

The following explanation of the effect, which is different from that suggested by White, occurred to me some time ago on theoretical grounds, and has since been confirmed by me experimentally. The effect depends on the fact (which is itself a consequence of theory) that when the pressure with which the bow is applied is less than a certain critical value proportionate to the rate of dissipation of energy from the string, the principal mode of vibration of the latter, in which the fundamental is dominant, is incapable of being maintained and passes over into one in which the octave is prominent. When the bow sets the string in vibration the instrument is strongly excited by sympathetic resonance, and the rate of dissipation of energy rapidly increases and continues to increase beyond the limit up to which the bow can maintain the string in the normal mode of vibration. The form of vibration of the string then alters into one in which the fundamental is feeble compared with the octave. Following this, the amplitude of vibration of the belly decreases, but this change lags behind that of the string to a considerable extent. When the rate of dissipation of energy again falls below the critical limit, the string begins to regain its original form of vibration with the dominant fundamental. This is accordingly followed, after an interval, by a fresh increase in the vibration of the belly, and the cycle then repeats itself indefinitely.

From *Nature* **97**, 362 (29 June 1916).

Lens proteins

More molecular opportunism

Russell F. Doolittle

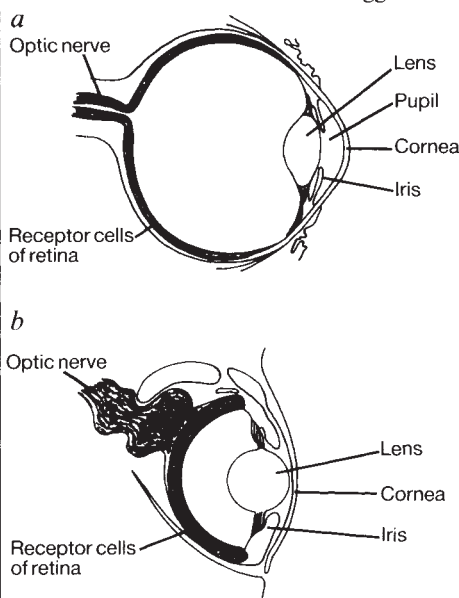
THE morphological resemblance of cephalopod eyes to those of vertebrates has long been considered a fascinating case of anatomical convergence (see figure), as it is well known that image-forming eyes have evolved independently on several occasions. In cephalopod and vertebrate eyes, for example, the cells in the retina are organized in completely different ways. It is known that the vertebrate eye's lens indulges in a kind of genetic piracy, and on page 86 of this issue¹, Tomarev and Zinovieva confirm that the cephalopod lens is guilty of a similar offence — the enslavement of an enzyme for a structural role. Lenses are found in other animal visual systems besides cephalopods and vertebrates, of course, including other molluscs (gastropods), spiders, polychaet worms, flatworms and even coelenterates². It will be of great interest to find out if those transparent structures have also pirated enzymes for structural purposes.

The bizarre story of the evolution of lens proteins began in 1982 with the report³ that α -crystallin, the predominant protein in the lenses of most vertebrate eyes, is similar to small heat-shock proteins. These heat-shock proteins are now known to be common among all eukaryotes and other types of structural proteins, including a schistosomal eggcase protein⁴, are also thought to be descended from them. The small heat-shock proteins could also be related to the ubiquitous 'prosome', an unusually stable ribonucleoprotein⁵. Further, the $\beta\gamma$ -crystallins, the other most common set of lens proteins, have structural features in common with a bacterial spore coat protein⁶, which implies that they too may be ancient.

The real surprise came when a routine computer search revealed that the amino-acid sequences of δ -crystallin from bird lenses and the enzyme argininosuccinate lyase are more than 55 per cent identical (see table). At about the same time, it was found that ϵ -crystallin, a protein of the lenses of birds and reptiles, is the active

glycolytic enzyme lactate dehydrogenase⁷. The ρ -crystallin from frog lenses was then shown⁸ to be about 50 per cent identical with human liver aldehyde reductase and the rat lens enzyme aldose reductase, and turtle τ -crystallin shows a remarkably close match to the enzyme enolase.

Even more surprising, a computer search of a published sequence fragment from a squid lens crystallin suggested⁹ that this protein might be the enzyme glutathione S-transferase. It is this suggestion



Comparison of *a*, vertebrate (human) and *b*, cephalopod (squid) eyes (from ref. 12).

that Tomarev and Zinovieva confirm in their paper in this issue¹. These authors cloned a family of squid lens proteins and show that the inferred amino-acid sequences of two of them are very similar to mammalian glutathione S-transferase.

What is to be made of these data? It is not news, of course, that old proteins can be duplicated and adapted to new functions. Well-known examples include the blood plasma protein haptoglobin, which has descended from the serine proteases, and the milk protein lactalbumin, which is derived from the enzyme lysozyme. The continuous expansion of the inventory of enzymes and other proteins is generally thought to follow from gene duplications and subsequent modification by base replacement. Certainly, the history of α - and $\beta\gamma$ -crystallins fits this scheme.

But the case of enzyme recruitment by the lens seems to be different. Piatigorsky *et al.* recently showed¹⁰ that the protein being over-expressed in the avian embryonic lens is argininosuccinate lyase

itself and not a duplicated congener, and the same case has just been made by Hendriks *et al.*¹¹ for lactate dehydrogenase and duck ϵ -crystallin. The same situation could be true of other enzyme-crystallin systems. The degree of sequence similarity observed for some of the systems seems lower than expected for single-copy active enzymes, however, given the phylogenetic situation of the species involved. Although an amphibian enzyme might be expected to be more than 75 per cent identical with its mammalian counterpart, the frog lens ρ -crystallin is only about 50 per cent identical with mammalian aldose or aldehyde reductases⁸. Frog and mammalian cytochromes *c* are more than 80 per cent identical, for example, and even their haemoglobin α -chains are more than 55 per cent identical. Similarly, a typical molluscan enzyme ought to be greater than 50 per cent identical in sequence to the corresponding mammalian enzyme, but the squid crystallins are only about 25 per cent identical with the mammalian glutathione S-transferase. The latter case is complicated in that this dimeric enzyme actually comprises a family of subunit types, as apparently do the squid lens proteins.

It seems unlikely that these enzymes are present in the lens for metabolic reasons. So why should they have a structural role in the lens? Crystallins are well known to be water-soluble proteins that can pack together efficiently to form very large molecular aggregates, and it could be that the only requirement for a structural lens protein is for a globular protein that can pack well and form transparent aggregates.

Or, as Piatigorsky *et al.* suggest¹⁰, maybe it has to do with an aspect of gene regulation that allows for tissue-specific expression of certain metabolic enzymes, and one that can be overcome in a tissue like the lens. Whatever it is, the independently evolved molluscan system has obviously hit on the same strategy. What has long been regarded as a model of anatomic convergence now also appears to involve a convergence of genetic enterprise. □

1. Tomarev, S. I. & Zinovieva, R. D. *Nature* **336**, 86–88 (1988).
2. Wells, M. *Lower Animals* (World University Library, New York, 1968).
3. Ingolia, T.D. & Craig, E.A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2360–2364 (1982).
4. de Jong, W.W. *et al. J. biol. Chem.* **263**, 5151–5149 (1988).
5. Martins de Sa, C. *et al. J. molec. Biol.* **187**, 479–493 (1986).
6. Wistow, G., Summers, L. & Blundell, T. *Nature* **315**, 771–773 (1985).
7. Wistow, G., Mulders, J.W.M. & de Jong, W.W. *Nature* **326**, 622–624 (1987).
8. Carper, D. *et al. FEBS Lett.* **220**, 209–213 (1987).
9. Wistow, G. & Piatigorsky, J. *Science* **236**, 1554–1556 (1987).
10. Piatigorsky, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 3479–3483 (1988).
11. Hendriks, W. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 7114–7118 (1988).
12. Purves, W.K. & Orians, G.H. *Life, the Science of Biology* (Grant, 1983).

Russell F. Doolittle is at the Center for Molecular Genetics, University of California, San Diego, La Jolla, California 92093, USA.

Enzymes corresponding to various types of lens crystallin

Enzyme	Crystallin	Occurrence
Argininosuccinate lyase	δ	Birds
Lactate dehydrogenase	ϵ	Birds Reptiles
Aldose reductase	ρ	Amphibians
Enolase	τ	Birds Reptiles Fish
Glutathione S-transferase	S	Squid

Alloy strength

Keeping dispersions fine

Robert W. Cahn

HIGH-TEMPERATURE alloys, especially those used in aero engines, have to serve at temperatures well above half their absolute melting point. Extremely fine dispersions of insoluble intermetallic phases that impede dislocation motion are essential to prevent failure, the only alternative being strengthening by ceramic fibres, a strategy that is still being developed. The fine dispersion must resist coarsening for a particular matrix/dispersoid combination to be useful. Ostwald ripening, the growth of larger particles at the expense of smaller ones by a process of solute diffusion through the matrix, is not important if the particles are very insoluble in the matrix. But an alternative process, the dragging of particles attached to the migrating boundaries of growing grains, can still force coarsening: the particles which are being dragged sweep up stationary particles from the body of a grain as the boundary moves past them. This process has not been analysed as much as it deserves, and hence three recent papers on the subject by Russell and Froes^{1,2} and by Hartland *et al.*³ deserve a special welcome.

Russell and Froes¹, who analyse the thermodynamic factors that favour the creation of a fine-particle dispersion in the first place, discuss the probable rate of Ostwald ripening. They also analyse the limiting rate at which a particle can keep up with a moving grain boundary; this depends both on the particle diameter and on its solubility in the matrix. The authors conclude that a drift rate of $1 \mu\text{m h}^{-1}$ can be maintained by particles of 1, 10 and 100 nm radius for low (10^{-4} atom %), intermediate (0.1 atom %) and high (10 atom %) solubilities, respectively. This is fast enough to lead to a potentially dangerous degree of sweeping, at least for the larger particles.

The analysis is dealt with in greater detail in Russell and Froes's second paper². The calculation is done on the assumption that the particles migrate by the process of diffusion along the particle/matrix interface, presumed to be faster than diffusion through the bulk of the particle. There is still uncertainty, dating back to a classic paper by Ashby and Centamore⁴, about whether crystalline particles can drift at anything like the rate of amorphous particles (such as silica), because crystalline particles tend to have a more ordered interface with the matrix, associated with slower diffusivities in the interface.

Russell and Froes compare experimental observations of various dispersoid systems with their calculations, with

special reference to the important case of Ti_3Al strengthened with dispersions of Er_2O_3 or Ce_2S_3 . It turns out that the greater resistance of the latter to coarsening can be explained in terms of the characteristics of the two dispersoids.

Hartland *et al.*³ study the special case of a population of empty or gas-filled pores drifting with a migrating boundary. This theme is older than the problem of particle drift: inclusion drift in a crystalline material was first observed for helium bubbles in copper⁵, and this led to the first theoretical treatment of bubble drift in a temperature gradient⁶ (without attachment to grain boundaries). The possible role of grain boundaries in dragging pores during the process of powder sintering and thereby hastening the elimination of porosity was pointed out several years ago⁷, but evidence of such dragging is very scant and uncertain.

Hartland *et al.*³ are the first to devise a theory of the dragging of pores by boundaries. Their special interest is the behaviour of fission-gas pores in uranium dioxide nuclear fuel. The authors analyse in rigorous detail the behaviour of various grain and pore geometries, including a range of porosity fractions and sizes, and for each situation calculate the retardation factor imposed by the presence of pores, compared to an identical grain geometry

without pores. One factor they take into account (this has never been considered before) is the influence of misorientation across a grain boundary; this affects the curvature of the interface of a trapped pore and thereby influences surface diffusion along that interface. It also turns out that pores retard grain boundary mobility more on grain edges than when at grain corners.

One variable not considered in this work is the effect of impurities dissolved in the matrix segregating to the pore/matrix interface; it has been shown experimentally, in metallic systems⁸, that such segregation can enhance surface diffusivity by factors of up to 10^4 . In practical terms, it would be interesting to know whether segregation of this kind is possible to the interface between dispersoid particles and the matrix. If so, it is conceivable that dispersoid mobility and thus dispersoid coarsening through drift of particles attached to grain boundaries could be considerably enhanced in the presence of particular impurities. □

1. Russell, K.C. & Froes, F.H. *J. Met.* **40**, 29–35 (1988).
2. Russell, K.C. & Froes, F.H. *Scripta metall.* **22**, 495–499 (1988).
3. Hartland, P., Crocker, A.G. & Tucker, M.O. *J. nucl. Mater.* **152**, 310–322 (1988).
4. Ashby, M.F. & Centamore, R.M.A. *Acta metall.* **15**, 1081–1092 (1968).
5. Barnes, R.S. & Mazey, D.J. *Proc. R. Soc. A* **275**, 47–57 (1963).
6. Shewmon, P.G. *Trans. Am. Inst. Min. Metall. Engrs* **230**, 1134–140 (1964).
7. Heuer, A.H. *J. Am. Ceram. Soc.* **62**, 317–318 (1979).
8. Rhead, G.E. *Surface Sci.* **47**, 207–221 (1975).

Robert W. Cahn is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3PZ, UK

Farming biotechnology

Regulating growth of animals

G. E. Lamming

LEADING international researchers took the opportunity at a recent meeting* to urge caution in making the assumption that biotechnology will necessarily improve farming practices by increasing growth rates of farm animals, resulting in 'super' strains. There is widespread fear that such techniques could exacerbate food surpluses and put small farmers out of business. Recent experiments indicate that the growth rate of farm species is not dramatically increased. It seems that centuries of genetic selection could already have produced individuals that cannot, or do not, respond markedly with increases in growth either following growth hormone gene insertion or parenteral treatment with recombinant growth hormone preparations.

* International Symposium on Biotechnology in Growth Regulation, AFRC Institute of Animal Physiology and Genetics Research (IAPGR), Babraham, Cambridge, 18–20 September 1988. Proceedings to be published by Butterworths.

Despite these potential drawbacks, it is clear that biotechnology can be effective in other ways. Treatment with recombinant bovine somatotrophin (rBST), for example, can consistently increase milk yield with improved feed-conversion efficiency and no significant changes in milk composition even in high-yielding cows. This hormone also results in leaner lamb and pig carcasses, thus providing healthier meat for human consumption. Its main physiological effect is that metabolism is influenced in favour of net protein gain.

There could, of course, be other explanations of why some genetically modified or growth-hormone-treated animals do not grow faster, and these problems were addressed at the meeting. In the aftermath of recent political action which resulted in a ban on the use of anabolic steroids for growth promotion of animals within the European Economic Community (EEC),

it is essential to ensure that EEC as well as national authorities do not legislate until research is complete and the results have emerged. Independent assessors advising licensing authorities now require information on the mode of action of the active ingredients in medicinal products as well as evidence of their effects on metabolism, and the latest data pertaining to these requirements were reported at the meeting.

Following recent interest in the structure of growth hormone in different species, attention is now being focused on characterizing growth-hormone receptors (M. J. Waters, University of Queensland). The receptor in the rabbit liver and bone is up-regulated by the level of growth hormone in circulation and has 80 per cent sequence homology with the human liver growth-hormone receptor as well as some similarity with the prolactin receptor. The next step is to determine the upstream and downstream control sequences.

There is on the other hand evidence of some heterogeneity of receptor types in individual animals with potential specificity for somatogenic, lactogenic and somatomammogenic effects (P. Gluckman, University of Auckland). The ontogeny varies with the tissue and no growth-hormone receptors have been isolated so far in the fetal liver, but in lambs a somatogenic receptor appears about 6 days after birth. Furthermore, there are both high- and low-affinity receptors and the level of high-affinity receptors, important in the control of growth, are probably insulin-mediated (Gluckman). Nutritional influences are important, with a high plane (generous nutrient intake) stimulating increased concentrations of high-affinity receptor sites. Of considerable interest in the agricultural context is that oestrogens increase both growth-hormone receptors and insulin-like growth factor-1 (IGF-1) level under a high nutritional plane but have little effect under a low plane.

Concerning the hypothalamic control of growth-hormone secretion, which involves both growth hormone releasing factor and somatostatin, it was suggested that acute feedback mechanisms operate via somatostatin whereas control of basal growth-hormone level is via IGF-1 (M. Scanlon, University of Wales, Cardiff). The pattern of growth-hormone administration may be an important factor governing the growth response (I.C.A.F. Robinson, National Institute for Medical Research, London) because at equivalent doses, pulsatile administration has a superior effect in hypophysectomized rats compared with constant infusion. There are sex differences in the normal pattern of growth-hormone secretion in rats; the low, flat mode characteristic of the female can be converted into the male episodic mode by intermittent somatostatin infusion, which results in extra weight, bone

growth and a higher growth-hormone level (Robinson).

High-plane nutrition during treatment and its influence on milk yield in dairy cows and carcass composition in lambs is beneficial (I. Hart, Cyanamid, New Jersey; J. Pell, IGAP, Hurley), and changes in both IGF-1 and IGF-2 in relation to milk yield are important. In response to fears that IGF-1 levels in milk from rBST-treated cows might be unduly elevated and thus affect new-born babies, it was reported that normal cow's milk contains large quantities of IGF-1 during the peri-parturient period. Human breast milk contains high levels of IGF-1 and there is an unexplained, substantial increase in IGF-1 in this milk at about 6–8 weeks post-partum (C. Prosser, IAPGR; R. Collier, Monsanto, St Louis). Fears of consumers at this point should thus be assuaged. Clinical interest in therapeutic properties of growth-hormone administration in relation to control of bone growth is supported by evidence from receptor-antibody studies that both proliferation and differentiation of cells controlling bone growth are IGF-1-dependent (Z. Hochberg, Rappaport Family Institute for Medical Research, Haifa).

The latest results of transgenic studies with both pigs and sheep were reviewed at the meeting. The efficiency of gene integration in pigs using the metallothionein promoter (0.9–1.42 per cent) and the percentage of expression has been disappointingly low in US Department of Agriculture trials (V. Pursel, Bettsville Agricultural Research Center, Maryland), with no relationship between copy number and growth-hormone level or indeed between growth-hormone level and growth rate. Eight of fourteen transgenic animals studied died before 90 days of age and there were significant problems of lethargy, muscle weakness, uncoordination, lack of libido and susceptibility to stress in these transgenic pigs. Similar results were reported from more limited studies in sheep (C. D. Nancarrow, CSIRO). Embryo manipulation as a route for transgenesis, perhaps using embryonic stem cells, is being investigated (E. J. C. Polge, Animal Biotechnology Cambridge Ltd). In addition, researchers at IAPGR and Ohio University have exploited a new construct consisting of a prolactin promoter and the bovine growth-hormone gene that can be inserted into pigs. Bovine growth-hormone levels in the circulation are low but induced expression can be achieved in a pulsed fashion with appropriate agonists or antagonists. All these data highlight the need for more precise control of the time, mode and level of gene expression. □

G. E. Lamming is professor of Animal Physiology at the University of Nottingham Faculty of Agricultural and Food Sciences, Loughborough, Leicestershire LE12 5RD, UK.

Daedalus

Bending the bugs

THE central problem of food preservation is killing the resident bacterial population of the food. Traditional methods like boiling and salting are very brutal; even pasteurization and gamma-radiation can alter the taste and character of the food. Now Daedalus has a new idea. He plans to give bacteria the 'bends'.

A diver gets the bends if he comes up to the surface too quickly. In the high pressure of the ocean depths, nitrogen from his air supply dissolves in his body fluids, and emerges as bubbles if the pressure is reduced too rapidly. In the bloodstream a bubble is a nuisance; in synovial or spinal fluid it can be very painful and dangerous. In a bacterium, says Daedalus, it would be fatal. The sudden expansion would blow the bacterium up like a tiny balloon and burst its cell wall.

So DREADCO biologists are subjecting selected bacterial cultures to sustained high gas pressures. When the organisms are saturated with gas, the pressure is suddenly released. Spherical cocci have little choice but to swell and burst; rod-like bacilli could perhaps accommodate their sudden flatulence by deforming into spheres, but should still be fatally disorganized by the experience. Spirilli would be disastrously unwound, like a Bourdon pressure gauge. The ideal gas for the process needs to be harmless and fairly soluble in water; nitrous oxide should be a better bet than nitrogen itself. Ten or twenty atmospheres should be sufficient to guarantee that every organism will absorb enough gas to form a bubble inside it when the pressure is released.

Any food product could be 'bent' in this way. Fluids like milk will simply froth up violently as the pressure comes off, and then collapse back to their original state; finished products like pies and sausages should first be sealed in a microporous packaging film permeable to gas but not bacteria. 'Bending' will cause no chemical damage at all, and may even enhance the texture of the denser foodstuffs. They could well expand permanently, acquiring a welcome lightness and fluffiness. Indeed, entirely new 'foamed foods' may emerge from the research, a range of nutritive generalizations of popcorn. The lightest imaginable puff-pastry, ethereally bouyant sponge-cakes, huge insubstantial foamed steaks melting in the mouth but disconcertingly pneumatic on the plate, expanded vegetables of thistledown rarefaction — all may be possible. Their guaranteed freedom from bacterial contamination, their subtlety and lightness of texture, and their unprecedentedly low calorie content per unit volume, should ensure them a wide welcome in the market-place. David Jones

Viral distemper now found in porpoises

SIR—Although thousands of seals (*Phoca vitulina*) have died in European seas since April 1988 from a morbillivirus infection, there have been no reports of morbillivirus infection in other marine mammals during the current epizootic. We now document distemper in two common porpoises (*Phocoena phocoena*) recently found dead on the coast of Northern Ireland.

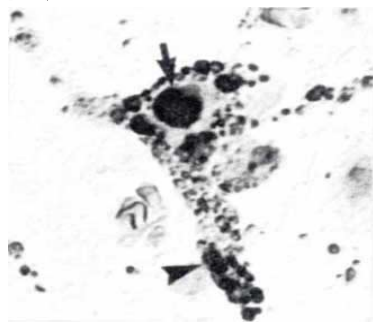
Necropsy reveals severe pneumonia characterized by proliferation of type II pneumonocytes, syncytia formation, inflammatory cell infiltration and necrosis of bronchial and bronchiolar epithelium. Acidophilic intracytoplasmic inclusion bodies highly specific for morbillivirus infection are seen in bronchial and bronchiolar epithelium. There is widespread degeneration and necrosis of neurons, gliosis and perivascular cuffing in the brain consistent with viral aetiology. Many neurons contain inclusion bodies. Inclusions are also seen in transitional epithelium of the renal pelvis and urinary bladder.

We examined paraffin sections of tissues from both porpoises for morbillivirus antigen by an immunoperoxidase technique. Measles antiserum from a patient with subacute sclerosing panencephalitis was used as the primary antibody. Tissues from a seal with distemper and a porpoise necropsied at our laboratory in 1987 were used as positive and negative controls respectively.

Morbillivirus antigen was detected in both animals. As shown in the figure, there is specific intracytoplasmic and intranuclear staining in neurons and astrocytes in the brain. Antigen is also seen in bronchial and bronchiolar epithelium and pneumonocytes in the lung and in transitional epithelium of the renal pelvis and urinary bladder. The microscopic changes in our porpoises are similar to those of canine distemper virus infection in other species including seals^{1,2}. Demonstration of morbillivirus antigen in diseased tissue confirms such a virus as the primary cause of the lesions.

This is the first report of morbillivirus

infection in cetaceans. We do not know how the porpoises contracted the infection. The pattern of spread of the seal epizootic is temporally and geographically similar to the annual migration of porpoises from the Baltic to the North Sea and south to the English Channel³, so it is possible that porpoises acted as vectors in the southerly



Immunoperoxidase staining of morbillivirus antigen in nucleus (arrow) and cytoplasm (arrowhead) of a neuron. $\times 650$.

spread of the seal epizootic. Porpoises are known to cross the Atlantic Ocean, so marine mammals along the American continent could be at risk.

Our findings may explain the declines in porpoise and dolphin populations in European waters in recent years. Retrospective serological examination of stored cetacean blood samples, and detailed post-mortem and serological examinations of cetaceans found ill or dead would help establish the timing and extent of morbillivirus infection in these species.

S. KENNEDY

J. A. SMYTH

P. F. CUSH

S. J. MCCULLOUGH

G. M. ALLAN

Veterinary Research Laboratories,
Stormont, Belfast BT4 3SD, UK

S. McQUAID

Department of Pathology,
Queen's University, Belfast BT7 1NN, UK

1. Kennedy, S. *et al.* *Nature* **335**, 404 (1988).

2. Dungworth, D.L. in *Pathology of Domestic Animals* Vol. 2 (eds Jubb, J.V.F., Kennedy, P.C. & Palmer, N.) Ch. 6 (Academic, London, 1985).

3. Matthews, L.H. *The Natural History of the Whale* 131 (Weidenfeld and Nicolson, London, 1978).

Did virus transfer from harp seals to common seals?

SIR—A morbillivirus related to canine distemper virus (CDV) is apparently responsible for the recent epidemic deaths of common (or harbour) seals, *Phoca vitulina* L., in the North Sea and the Baltic¹. Osterhaus *et al.*², investigating the virus in seal populations, have reported finding no CDV-neutralizing antibodies in samples collected from eight different seal species before the present outbreak of the disease. These include four of the five phocine genera (*Phoca*, *Pusa*, *Histrio-*

phoca and *Halichoerus*).

Pagophilus, the fifth phocine genus, ought to be investigated as well, since the harp seal, *Pagophilus groenlandicus* Erleben, is perhaps the most likely source of the virus.

Vast numbers of harp seals live in the White Sea and elsewhere in the far north, but last year there was a massive irruption of them in to the North Sea. This could have enabled the virus to jump from harp seals, in which the disease may long have

been endemic allowing acquired immunity to the virus to develop, to common seals, which had not been previously exposed to the virus. The disastrous results would be similar to what happened when the myxomatosis virus, endemic in American cottontail rabbits (*Sylvilagus* spp.) as a mild disease, was introduced into the European rabbit, *Oryctolagus cuniculus* (L.) in the 1950s, where it was highly lethal.

This hypothesis could be tested by attempting to infect harp seals with virus taken from diseased common seals, when it should produce only relatively mild symptoms. And CDV-neutralizing antibodies would be expected to be found in samples taken from harp seals within their normal range both now and, if any are available, from previous years.

C. B. GOODHART

Gonville and Caius College,
Cambridge CB2 1TA, UK

1. Osterhaus, A.D.M.E. & Vedder, E.J. *Nature* **335**, 20 (1988).

2. Osterhaus, A.D.M.E. *et al.* *Nature* **335**, 403 (1988).

Is bacterial evolution random or selective?

SIR—J. Cairns *et al.* (*Nature* **335**, 142–145; 1988) raise the question of whether bacterial evolution always proceeds by a random process or if bacteria have the ability to sense specific needs and direct mutational events to satisfy those needs. The present generation of microbial geneticists, having been taught in their formative years that mutations are random events and that selections serve only to identify those rare individuals which contain the appropriate mutation, will take comfort in the reconfirmation of the Luria and Delbrück results by Cairns *et al.*. But the idea that bacteria can specifically sense metabolic needs during a selection and respond by directed mutagenesis will no doubt be met with much scepticism, for the questions raised involve an aspect of biology we know little about, that is, adaptation and macromolecular synthesis during starvation conditions. Although the actual experimental data presented are sketchy and the details of the experimental procedures are lacking, one is left with the feeling that the phenomenon reported by Cairns *et al.* is real and warrants further investigation.

I also have been investigating the events that lead to the appearance of mutant colonies on agar plates during a non-lethal selection. This selection demands that cells grow on a carbon source whose molecules are too large to easily cross the outer membrane. I obtained mutants which continued to appear for many days after plating. These mutants have mutations that predominantly alter one of the outer membrane porin proteins (OmpF).

When the *ompF* gene is deleted, alterations to a similar outer membrane porin, *OmpC*, are obtained. Why these *ompC* mutations are not obtained when the cell is *ompF*⁺ remains a mystery since the alterations in both *ompF* and *ompC* are analogous at the DNA and protein levels and reconstruction experiments show the *ompC* mutations are sufficient to allow colony formation in a lawn of *ompF*⁺ cells (S.A.B. *et al.* *J. molec. Biol.* in the press). I believe that we are seeing a form of directed mutagenesis in this selection, though we have no data on how this directed response might occur. One explanation is that the *ompF* gene, due to its chromosomal position, is preferentially mutated when the cell senses a need to increase its outer membrane permeability during starvation. As pointed out by F. Stahl (*Nature* **335**, 112–113; 1988) one could envision a number of possible molecular mechanisms to account for how cells might, under certain circumstances, direct or influence mutational events. Further tests of the notion of a directed mutagenic response during non-lethal selections are needed. We owe a debt of gratitude to the Cairns group for pointing out the limitation of the evolutionary dogma formulated in the 1940s and 1950s and for helping to pave the way for studies that re-address the question of how bacteria evolve and respond to evolutionary stress.

SPENCER A. BENSON

Department of Biology,
Princeton University, Princeton,
New Jersey 08544, USA

SIR—J. Cairns *et al.* (*Nature* **335**, 142–145; 1988) have presented experimental evidence and circumstantial information which they claim supports the occurrence of “directed mutation” in bacteria in stationary (non-growing) phase. Their crucial evidence for this assertion came from an experiment in which *Lac*[−] *Escherichia coli* were plated on a medium (M9) which supposedly kept them in stationary phase. After various times in this condition, they were exposed to a lactose-containing medium which supposedly also prevented growth, and on which revertants to *Lac*⁺ genotype could be scored by colony growth. The authors argue that if the mutations producing revertants were occurring before lactose was added, the number of colonies growing on the lactose medium should have increased with the length of time spent in the minimal medium (M9) before the lactose was added. If the mutations did not occur until the lactose was added, then the number of colonies using lactose should be a function only of the time since lactose addition. The latter was claimed to be the case, apparently supporting the occurrence of directed mutation rather than the considerably more conventional random mutation model.

It is crucial to the argument that the detection of mutations occurring before and after exposure to lactose was the same; Cairns *et al.* imply that all mutations occurring in both conditions were detected by the methods used. However, the logic of this relies on a number of unwarranted assumptions underlying the interpretation of the experiment, the most important of which is that neither death nor growth was occurring.

During culture on the M9 plates before lactose exposure, death of some revertants may have occurred. Death of revertants would tend to lead to a negative correlation between time spent on the minimal medium and subsequent absolute number of colonies growing on exposure to lactose. This negative correlation could act to reduce any positive correlation arising from the accumulation of mutants. If death outweighed the accumulation of mutants then the net number of mutants would decrease as a function of time. This is what Fig. 3 of Cairns *et al.* shows; the growth curves in the presence of lactose do not superimpose, but start off at a lower level in the late-exposed cultures. The authors themselves note that the number of *Lac*⁺ colonies detected actually goes down with time on M9 plates, supporting this argument. It is not necessary for the argument that the death rate of *Lac*⁺ and *Lac*[−] cells differed; any death of *Lac*⁺ cells would lead to an underestimate of mutation rate. The estimate of random mutation rate before lactose exposure may therefore have been too low.

To overcome the death problem one could plot the number of *Lac*⁺ colonies as a proportion of the total number of cells still alive, rather than as the absolute number of colonies which Cairns *et al.* measured. Even then, however, there would be a further difficulty if cells of the two genotypes differed in survival probability in the absence of lactose.

The second major difficulty is that the mutation rate after exposure to lactose may well have been elevated by growth. Cairns *et al.* themselves note that there would be a problem if the *Lac*[−] cells were better able to grow in the presence of lactose than on M9 alone, either because the mutant was leaky or because the lactose was impure. If lactose medium supported slow growth, and if the resulting extra DNA replication allowed the accumulation of more mutations than in culture on M9, then the number of mutant colonies detected would be a function of time since the lactose was added for this reason alone, without any need for the occurrence of directed mutation. Cairns *et al.* argue against this possibility by showing that a different mutation (conferring valine resistance) does not accumulate during the time that the colonies are exposed to lactose; presumably, then, they are not accumu-

lating merely random mutations as a function of growth. But this argument is flawed. It crucially assumes that the mutation conferring valine resistance is selectively neutral in the absence of valine. Such an assumption is unwarranted, because the mutation may reduce survival rate under these culture conditions. This, in addition to random cell death, could explain the finding that the number of Val(R) colonies actually decreased as a function of time spent in the lactose medium before the valine challenge. We can conclude, then, that there may be selection against Val(R) in the absence of valine, and the logic of the crucial control experiment collapses.

The design of the experiment was complex, and the interpretation hinges critically on the assumptions that there was no death and no replication. It would seem preferable to search for a design that avoids these assumptions since it is hard to see how they could be directly validated.

In a second experiment Cairns *et al.* describe further evidence based on reversions to *Lac*⁺ in an *E. coli* strain with the *lacZ* region under the control of an arabinose positive regulator, but rendered *Lac*[−] by the presence of a bacteriophage Mu DNA insertion between them. Deletion of this sequence causes reversion, and apparently occurred at a higher rate on minimal medium to which lactose and arabinose had been added than in rich cultures deficient in these two sugars. However, there was no control experiment to measure the deletion rate of this bacteriophage sequence from other parts of the *E. coli* genome under these conditions, so that it is not clear if the mutations were genuinely directed. Furthermore, the interpretation relies on the untested assumption that the survival of revertants in the absence of lactose and arabinose was as good as that of the original strain; the point that we raised earlier in connection with the valine-resistance study.

Finally, even if the Cairns *et al.* interpretation of the *Lac*(*Ara*)⁺ experiment was correct, it is in their own words “rather a special case”, because it appears to involve a specific mechanism for the movement of a foreign DNA fragment, rather than a mutation in the ordinarily accepted sense of the word. We submit that conceptual disarray might result from trying to solve the problem of random versus directed mutation by studying the complex rules of warfare between bacteria and their parasites.

LINDA PARTRIDGE

Department of Zoology,
University of Edinburgh,
Edinburgh EH9 3JT, UK

MICHAEL J. MORGAN

Department of Psychology,
University College London,
London WC1E 6BT, UK

The Challenger challenger

Daniel S. Greenberg

"What Do You Care What Other People Think?" Further Adventures of a Curious Character. By Richard P. Feynman as told to Ralph Leighton. W.W. Norton: 1988. Pp.224. \$17.95. To be published in Britain by Unwin Hyman in February 1989.

BEHIND the tinny title of these collected writings, there's a treasure: the late Richard Feynman's enchanting, 110-page account of his maverick burrowings into the empire of the National Aeronautics and Space Administration. It is a wry masterpiece of inquiry into organization, published under a bland chapter heading, "Mr Feynman Goes to Washington". The contents of the rest of the book range from mere filler of little consequence, to Feynman's restrained but sorrowful recollections of his courtship of his childhood sweetheart and her death from cancer shortly after their marriage during the Second World War.

Feynman, who died last February, was indeed a curious character. A Nobel laureate in physics, he was revered as a teacher and renowned as a prankster. (An expert lock picker, he ridiculed the loose security during his service with the wartime atom bomb project by opening supposedly secure safes and leaving notes signed "Feynman the safecracker". This episode and others are related in an earlier work, *"Surely You're Joking, Mr. Feynman!"* *Adventures of a Curious Character*, published by Norton in 1985).

Feynman relished masking his perspicacity with a yokelish manner and disavowal of understanding. He also shunned the committee circuit. Thus, "Not me", he said, when a former student, William Graham, then acting head of NASA, invited him to take time off from the California Institute of Technology to serve on the presidential commission investigating the destruction of the *Challenger* shuttle. "I have a principle of not going anywhere near Washington or having anything to do with government", Feynman explains to his readers, "so my immediate reaction was — how am I gonna get out of this?"

His wife, however, saw the matter differently. "If you don't do it", she said, "there will be 12 people, all in a group, going round from place to place together. But if you join the commission, there will be 11 people — all in a group, going round from place to place together — while the 12th one runs around all over the place, checking all kinds of unusual things".

Feynman signed on. But before leaving for Washington, he took two important steps. Unlike most of his fellow commissioners, he cancelled all other obligations, for six months, and he enlisted experts at

the nearby Jet Propulsion Laboratory for a tailor-made quick course in space engineering. As his wife had predicted, he was soon running about all over NASA, checking all kinds of unusual things.

From the start, Feynman grated on the commission chairman, William Rogers, a former Secretary of State whose lawyerly devotion to procedural orderliness clashed sharply with Feynman's style of sniffing around for interesting information. "I was obviously quite a pain in the ass for Mr. Rogers", he acknowledges. Eager to plunge in, Feynman was dismayed by the leisurely pace of the *Challenger* inquiry: "Although it *looked* like we were doing something every day in Washington, we were, in reality, sitting around doing nothing most of the time".

Rogers, more eager to salvage NASA's reputation than to embarrass the agency, dismissed proposals by Feynman and others for an exhaustive, independent examination along the lines of an inquest into a commercial aviation disaster. An alarm sounded for Feynman when Rogers arranged for NASA to brief the commission at the Kennedy Space Center: "I get this picture of the czarina coming to a Potemkin village: everything is all arranged; they show us how the rocket looks and how they put it together. It's not the way to find out how things *really* are".

Told that five days free of commission

business would elapse before the visit to the Space Center, Feynman arranged through Graham — then still at the helm of NASA — to meet NASA technical specialists. These discussions brought out previously undisclosed difficulties with the so-called O-rings — rubber seals designed to cover joints in the booster rockets. Feynman was shown a NASA report describing the O-ring problem as "most critical", but the report concluded that "it is safe to continue flying existing design as long as all joints are leak checked. . . .". To Feynman, the report was preposterous. "I was struck by the contradiction", he writes. "If it's 'most critical', how could it be 'safe to continue flying'? What's the *logic* of this?". He had thus made the first important breakthrough of the *Challenger* inquiry: errant management, rather than failed technology, became a prime suspect.

The breakthrough widened as data were brought forth on the stiffening effects of low temperature on the O-ring material. An employee of Morton Thiokol, manufacturer of the boosters, volunteered to the commission that NASA had been warned to abort the mission if the temperature fell below 53°F. On the morning of the launch, it was 29°F. Shortly afterwards, at a televised hearing of the commission, Feynman became a national celebrity when he dropped a chunk of O-ring material into a glass of ice water, and, on retrieving it, showed that it had been chilled to a stiff state.

Feynman then looked into other mysteries of NASA's management of shuttle engineering, including the danger ratings of high-frequency vibrations from the craft's immensely powerful engines. He found that NASA fixed the problems where it could, but more or less tried to

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Feynman — sniffing around for information.

muddle through where it couldn't, and came away "with the definite impression that I had found the same game as with the seals: management reducing criteria and accepting more and more errors that weren't designed into the device, while the engineers are screaming from below, 'HELP!' and 'This is a RED alert!'".

Feynman and several of his colleagues on the commission soon recognized what they were up against: NASA management, scrambling to survive the catastrophe, had, in Washington bureaucratic parlance, reconfigured itself into the CYA (cover your ass) mode. The giant agency, strong in public relations, aimed to ride out the storm of national anguish and official inquiry, and emerge merely as the victim of technological happenstance. It might possibly have succeeded but for Feynman's irrepressible proings, which led to the heart of the matter: slovenly management that permitted wishful thinking to override basic engineering principles.

Feynman's final analysis of the malady of NASA should be mandatory reading for the managers of complex organizations, and in particular ageing complex organizations. The Apollo Moon landing, like the Manhattan Project of the Second World War, had succeeded, he concludes, because all hands shared the same goal and "everybody's interested in everybody's problems". After Apollo, NASA's survival depended on convincing Congress that a new project, the shuttle, would be a high-performing, economical means for reaping the benefits of space. The managerial salesmen thus put on a hard sell that could not tolerate cautions and doubts from engineers at the working levels.

"Well", writes Feynman, "the guys who are trying to get Congress to okay their projects don't want to hear such talk. It's better if they don't hear, so they can be more 'honest' — they don't want to be in the position of lying to Congress! So, pretty soon the attitudes begin to change: information from the bottom which is disagreeable — 'We're having a problem with the seals; we should fix it before we fly again' — is suppressed by big cheeses and middle managers who say, 'If you tell me about the seals problem, we'll have to ground the shuttle and fix it.' Or, 'No, no, keep on flying, because otherwise, it'll look bad,' or, 'Don't tell me. I don't want to hear about it'".

His final words on the *Challenger* disaster merit respectful consideration on every one of the issues of technology and public policy that vex modern society: "For a successful technology, reality must take precedence over public relations, for Nature cannot be fooled". □

Daniel S. Greenberg is Editor and Publisher of Science & Government Report, 3736 Kanawha Street NW, Washington, DC 20015, USA.

Windows on heaven

David H. DeVorkin

Astronomical Centers of the World. By Kevin Krisciunas. Cambridge University Press: 1988. Pp. 320. £17.50, \$24.95.

A SURVEY of the anatomy and personality of astronomical observatories during any age reveals much about the astronomical enterprise of that time, as well as the society that fostered it. Developing from visual sighting devices limited to detecting positions and motions, to modern optical, radio and space observatories, over the centuries the compass of astronomy has extended to include all known regions of the electromagnetic spectrum, limited only by available technology and the Earth's atmosphere. At some times, astronomers were themselves the technological innovators, but most typically they benefited from devices created for other purposes. At some times, astronomers rapidly grasped the new opportunities provided, and at others, they balked.

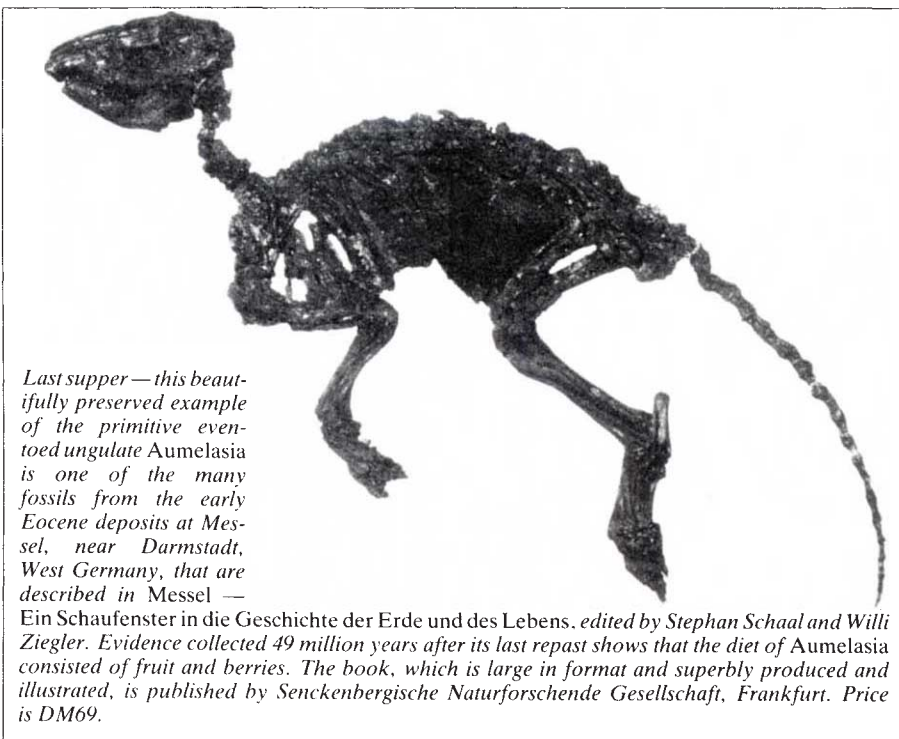
One might expect that a history of astronomical centres would contemplate these matters. In a discipline as diverse, complex and popular as astronomy, there is much fertile territory to dig over. Why were some technologies assimilated more rapidly than others? What social drives promoted the study of astronomy, and led to funding for its instruments and institutions? And who were the builders: the instrument makers and technologists who made it possible?

Astronomical Centers of the World addresses these matters, but only very

briefly. In what is essentially an extended chronological narrative highlighting about ten of the most illustrious observatories and institutions of the past two millennia, an ocean of names, dates and anecdotal detail washes over questions of how and why each of these venerable centres was built and to what purpose. There is an attempt at an overarching theme: in the preface the author contends that "at a given time there was an astronomical center whose equipment and staff were such that crucial, path-breaking work could have been done nowhere else. Or in some cases it could have been done somewhere else, but was not". But I have to admit that I am not sure what this means.

There is pleasant diversion to be found in this broad-brush digest. For me, the early chapters on the Alexandrian library and Ulugh Beg were interesting, possibly because I know little about their histories. Later chapters, however, on progressively more recent institutions from Tycho's Hven, the Paris and Greenwich observatories, Pulkovo, Harvard, Lick, Yerkes, Mount Wilson and Palomar left me less and less satisfied as the author moved into realms with which I am familiar. The least compelling accounts are those of comparatively new (post-1945) institutions, including the ones in Hawaii that Mr Krisciunas is intimately acquainted with. My mixed reactions thus suggest that this book will provide a few agreeable hours of reading for those mildly curious about astronomical history, but who are not sufficiently involved in the subject to demand fresh, critical analysis. □

David H. DeVorkin is Curator of Astronomy at the National Air and Space Museum, Smithsonian Institution, Washington, DC 20560, USA.



Last supper—this beautifully preserved example of the primitive even-toed ungulate Aumelasia is one of the many fossils from the early Eocene deposits at Messel, near Darmstadt, West Germany, that are described in Messel —

Ein Schaufenster in die Geschichte der Erde und des Lebens, edited by Stephan Schaal and Willi Ziegler. Evidence collected 49 million years after its last repast shows that the diet of Aumelasia consisted of fruit and berries. The book, which is large in format and superbly produced and illustrated, is published by Senckenbergische Naturforschende Gesellschaft, Frankfurt. Price is DM69.

A proliferation of plants

Peter D. Moore

Vegetation Mapping. Edited by A.W. Kùchler and I.S. Zonneveld. *Kluwer*: 1988. Pp. 635. Dfl 500, \$250, £154.

Vegetation Ecology of Central Europe, 4th edn. By Heinz Ellenberg. Translated by Gordon K. Strutt. *Cambridge University Press*: 1988. Pp. 731. £75, \$125.

North American Terrestrial Vegetation. Edited by Michael G. Barbour and William Dwight Billings. *Cambridge University Press*: 1988. Pp. 434. £45, \$49.50.

Vegetation of the Soviet Polar Deserts. By V.D. Aleksandrova. Translated by D. Löve. *Cambridge University Press*: 1988. Pp. 228. £30, \$49.50.

THERE are few terms that have caused more controversy among botanists than 'vegetation'. Vegetation may unite plants, but it certainly divides plant scientists. The main debate is to do with whether plant species behave in an independent, individualistic way, each distributed according to its own particular environmental limits of tolerance, or whether a process of co-evolution has resulted in the development of distinct units of vegetation, each unit being made up of a co-adapted assemblage of mutually tolerant, or even dependent, plant species. Generally, botanists from continental Europe have adopted a community approach to vegetation, whereas the English-speaking world has used the individualistic concept. One thing, however, unites the students of vegetation and that is the need to describe and map. For this purpose, vegetation units, whether real or contrived, are clearly necessary.

Kùchler and Zonneveld have put together a detailed critique of mapping methods and their theoretical bases. As they point out, there are two main approaches to the problem, largely determined by the scale at which vegetation is studied. At a local level, floristic analysis of vegetation is the most reasonable approach, though this raises further questions of demarcation and classification. At a global level, methods based on taxonomy are inappropriate and it is usual to turn to the gross morphology of vegetation as a basis for classification. The life form (or perhaps better, growth form) of the component plants is the basis of this approach, leading to an initial subdivision into forest, scrub, grassland and so on.

The book is broken up into 40 short chapters dealing with various aspects of vegetation classification and mapping, but it is still not easy to gain access to information. A discussion of floristic analysis, for example, precedes the account of field

survey. The approach to survey work, incidentally, has a strongly 'continental' flavour with floristic tabulations being preferred to computer analyses based on similarity indices. There is also, very reasonably, a strong emphasis on remote sensing; indeed, Zonneveld claims that "in the 20th century it is ill-advised to do vegetation surveys not using aerial photo-interpretation". It is in this area that the book is at its strongest.

Ellenberg is a name which has come to dominate vegetation science in mainland Europe and a translation of the fourth edition of his major work will be welcomed by all English-speaking plant ecologists. Ellenberg has succeeded in developing a phytosociology in which the ecology and physiology of the individual is not forgotten. As Peter Grubb puts it in his introduction, "Despite his phytosociological training and proper concern for communities, his account is based squarely on individual species and small groups of species". Ellenberg's approach contains a strong physiological element; indicator species are given scores with reference to their requirements for or tolerance of certain environmental factors (light, temperature, continentality, moisture, alkalinity and nitrogen) and hence ecological groups can be isolated and can assist in the description and ecological interpretation of vegetation patterns. His book is an immense compendium of the results of both field and laboratory work; both experimental and descriptive data are included, and the whole is set in its human and historical context, thus providing a key to the understanding of central European vegetation.

A particularly valuable feature of Ellenberg's work is his inclusion of the vegetation of disturbed habitats. In Europe one is far more likely to encounter arable fields and managed grasslands than natural vegetation, yet very few descriptive accounts provide information about their floristic composition. The part of the book dealing with these vegetation types will prove of value not only to ecologists but to agriculturalists, agronomists and even archaeologists trying to interpret the weed communities of the past.

North America is much larger than central Europe and the differences in scale demand a different approach to vegetation description, as is demonstrated by the collection of articles in Barbour and Billings's book. When dealing with a continent in which both tundra and tropics are represented, the detailed floristic accounts that form the basis of Ellenberg's work are not appropriate. The generally lower intensity and shorter history of advanced agricultural and industrial cultures in North America has also left the vegetation less fragmented and less intensively managed than that of Europe. It is reasonable, therefore, that this descrip-

tive account should confine itself largely to the natural and semi-natural vegetation types.

The initial subdivision of the vegetation is based on physiognomy and geography — Rocky Mountain forests, Pacific north-west forests, Intermountain deserts, southeastern coastal plain vegetation and so on. Within each of these units, further floristic subdivision is possible, sometimes using multivariate, numerical techniques, such as those applied by Ritchie in analysing the communities of the Yukon. The concept of vegetation as a continuum is inherent in many of the contributions, and the Whittaker system of positioning 'communities' along environmental axes, such as elevation and moisture gradients, is used to represent vegetation relationships in the Sierra Nevada, the Rockies and, of course, the Smoky Mountains where Whittaker first developed the method. The difference in approach to that found in Ellenberg's book in part reflects the more continuous nature of the relatively undisturbed vegetation of America, but it also reveals a difference in attitude. The North American approach looks upon communities as convenient (and somewhat artificial) abstractions, whereas in continental Europe the community is regarded as an evolved entity requiring classification in much the same way as an organism. ▶

ADVERTISEMENT

The ANS Journals

Your number one source for current information about
NUCLEAR ENERGY

NUCLEAR NEWS

NUCLEAR SCIENCE & ENGINEERING

NUCLEAR TECHNOLOGY

TRANSACTIONS

REMOTE SYSTEMS TECHNOLOGY PROCEEDINGS

FUSION TECHNOLOGY

For free catalog contact:

AMERICAN NUCLEAR SOCIETY

555 N. Kensington Avenue
La Grange Park, Illinois 60525 USA
312-352-6611 Telex: 4972673

Unlike Ellenberg, however, Barbour and Billings make little attempt to explain vegetation in terms of the individual physiology of species, though the account of the temperate deciduous forests does begin to move in this direction as the light requirements of the ground flora are considered. Here also the ecosystem concept is incorporated to some extent, and productivity and biomass data are used in the comparison of the various types of deciduous forest.

In the Soviet Union, the community concept (embodied in the 'phytocoenose') is alive and well, mainly because of the influence of Sukachev of the 'Leningrad School'. Something akin to Ellenberg's concept of ecological groups emerged from the opposing 'Moscow School', but has not been strongly influential in Soviet geobotany. Aleksandrova's account of the vegetation of the Soviet polar deserts has, therefore, a distinctly 'continental' ring to it, though the ecological group idea does emerge within the detailed analyses of the plant surveys described in the book. Also apparent is the use of global distribution maps which provide a biogeographical base for the establishment of component groups within communities.

One should emphasize that Aleksandrova's book is concerned solely with the 'polar deserts' and not the tundra in general. It includes only those areas of land which have their major portions covered by ice and which are characterized by a summer mean temperature not exceeding 2 °C. Under these conditions, the term vegetation can be used only loosely and lower plants (bryophytes and lichens) form an important part of the plant cover. This descriptive account has a strongly ecological emphasis, with detailed information on spatial patterns in relation to variations in soil and microclimate.

These four books, though similar in intent, demonstrate that the old conceptual divide still exists in vegetation science. The Atlantic Ocean is evidently more than a biogeographical barrier, it still restricts intellectual flow. In Britain we are awaiting the outcome of a national survey of vegetation which will result in a new descriptive account of the country's plant assemblages. Perhaps some British compromise will provide a bridge between the Old and New Worlds. □

Peter D. Moore is in the Division of Biosphere Sciences, King's College London, Campden Hill Road, London W8 7AH, UK.

All in a lifetime

David J. Thompson

Reproductive Success: Studies of Individual Variation in Contrasting Breeding Systems. Edited by T.H. Clutton-Brock. University of Chicago Press: 1988. Pp. 538. Hbk \$75, £59.95; pbk \$29.50, £23.50.

UNTIL recently, most attempts to compare variation in breeding success between animals in natural populations relied upon data collected over part of their lifespans and often from individuals whose age was not known. Such cross-sectional data, though illuminating in lots of ways, were not especially useful in providing estimates of reproductive success over the entire lifespan. For example, attempts to show that reproductive success changes with age using short-term data can easily be confused if there is a consistent association between breeding success and longevity.

In order to circumvent internal correlations of this nature, it is necessary to undertake more painstaking longitudinal studies in which the reproductive success of recognizable individuals is measured over successive breeding attempts, preferably over the whole lifespan of the individuals concerned. Reports of 25 such studies form the bulk of this book. There are five chapters on insects, one on amphibians, six on mammals and the inevitable majority (13) on birds, including

two contributions on great tits (both excellent).

The authors were invited to address four principal questions: "How widely does breeding success vary between individuals of each sex? How much of the variance in success is contributed by the different components of breeding success (survival to breeding age, reproductive life span, fecundity or mating success, and offspring survival)? To what extent does reproductive success change with age? And what environmental, phenotypic, developmental, or genetic factors affect breeding success in each sex?". The extent to which they have been able to answer these questions depends largely on the organism on which they have worked and whether their study was designed initially with lifetime reproductive success in mind.

The contributors of the data chapters have managed to stick to these common themes remarkably well. This is all the more remarkable given that some authors have very traditional population ecology backgrounds (for example Ollason and Dunnet, Thomas and Coulson, Newton) while others can be said to be principally behavioural ecologists (Partridge, Vehrencamp *et al.*, Fincke). Clearly, Clutton-Brock and his referees have done a good job. Other common themes are the enthusiasm with which many of the contributors stress the suitability of their animals for the study of lifetime reproductive success (Smith, Le Boeuf and Reiter), and how data have been ingeniously acquired

from less obliging subjects such as the Serengeti lions.

In addition to the compilations of data, there are three general chapters. In the first, David Brown discusses a method of partitioning variance in lifetime reproductive success and compares it with what has become the conventional method of Arnold and Wade. The ways in which these methods have been applied come in for further rigorous examination in the second general chapter, where Alan Grafen stresses the distinction between adaptation and selection in progress. An adaptation is a character of an organism which clearly serves a purpose and has been governed by natural selection in the past. Selection in progress is concerned with current changes in gene frequency of genetically different individuals. Put simply, the methods of Wade, Arnold and Lande for the most part detect selection in progress, whereas most students of lifetime reproductive success, including the majority of the authors of papers in this volume, are really dealing with the study of adaptation. Thus care must be taken in choosing methods of analysis.

Grafen's contribution is typically thought-provoking in other ways. Lest anyone gets too carried away in the euphoria of discovering longitudinal studies, he includes an excellent section on the disadvantages of the approach and on what cannot be inferred from such data. This discussion centres on why lifetime reproductive success is not equivalent to darwinian fitness. The final general chapter is left to the editor. Armed with the proofs of the data chapters, Clutton-Brock answers with great clarity some of the questions posed in his introduction.

There is only one mildly disappointing aspect about the book; it seems to have taken a long time from first requests for manuscripts to publication. Some chapters do not contain any references to work after 1985, even when relevant papers were published in 1986 and early 1987. Indeed, rather too many papers in the reference list are "in press, 1987" but were actually published early that year. Others are described as "Forthcoming 1987" or just "n.d."!

But nothing so trivial can detract from what is a notable book in the marriage between population ecology and behavioural ecology. It is essential reading for serious students in these disciplines, for it brings home the importance of detailed studies of individuals within populations. Moreover some marvellously inventive science is described — it is difficult to think of another collection where I have enjoyed the methods sections almost as much as the results. □

David J. Thompson is a Lecturer in the Department of Environmental and Evolutionary Biology, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK.

New perspectives on forest decline

L. W. Blank, T. M. Roberts and R. A. Skeffington

Prophecies that 'new-type forest damage' would end in Waldsterben (forest death) have not come to pass. Greater realism now characterizes research on the subject.

FORESTS are part of the heritage and folklore of all European countries, but their emotional appeal is probably strongest in Germany. In 1981, public attention was first drawn to the deterioration of the nation's forests by a series of articles in the magazine *Der Spiegel*, which reported that the forests were being damaged on an unprecedented scale. Such reports in the general press as well as in scientific journals triggered a lively and often emotional controversy over the extent and causes of the *neuartige Waldschäden* ('new-type forest damage'). The debate, however, suffered from a lack of reliable information on almost every aspect of the problem, and the opinions expressed were often characterized by preconceived ideas rather than fact. Principally, it was widely accepted that 'acid rain' (used as a synonym for air pollution in general) was the culprit. This view influenced political decision-making (for example, air-pollution abatement legislation was introduced rapidly) and dominated research on the problem.

Forest decline is no longer an issue confined to Germany. Widespread damage has been reported from most Central European countries and from the north-eastern United States, and has stimulated a great deal of research into its possible causes. The greatest contribution has come from West Germany, where large-scale funding became available from 1983. Many of these research projects are now almost complete. The results have greatly improved our understanding of the problem, as shown in reports of the West German government's Advisory Council on Forest Decline/Air Pollution¹ and in various reviews published since 1985 (refs 2–6). Also, other European countries have been carrying out forest health surveys for several years, allowing comparisons of damage to be made. This new flow of information has led to changes in our perception of forest decline, and here we discuss the developments of the past few years and consider their implications.

The first development has been the improved definition of damage, and the improved data on the geographical distribution and severity of decline. Damage to forests is not a new phenomenon and local problems have been reported ever since there have been reliable records^{7,8}. Such local events were generally traced back to specific causes (extreme climatic conditions, insect attack or local pollution, for

example). In contrast, recent forest damage in West Germany has affected a range of species over a broad area.

What is now generally called new-type forest damage became evident in the early 1980s, when Norway spruce (*Picea abies* L. Karst) was found to show damage symptoms such as needle discoloration and



Thin on top — spruce in the Black Forest.

crown thinning^{1,9}. Earlier local damage of silver fir (*Abies alba* Mill.) had not caused much concern, as this species has a well-documented record of diseases^{10,11}. But spruce decline did because of the species' wide distribution (spruce accounts for 39% of West Germany's forests) and its commercial importance as the 'bread-tree' of forestry. Damage to broad-leaved species, notably beech (*Fagus sylvatica* L.) and oak (*Quercus* spp.), which respectively make up 17% and 8% of forests, has progressively increased since 1984 and also became a matter of concern.

In 1984, the West German authorities

initiated standardized, annual and nationwide Forest Damage Surveys in order to give "a representative general account of the present state of the forests" (ref. 12). Two preliminary surveys were carried out in 1982 and 1983, but these cannot be directly compared with subsequent surveys because of differences in methodology⁵. All surveys after 1984 have been based on the visual assessment of trees at standardized sampling plots defined by a grid system. The damage classification is based on the estimated percentage of lost foliage compared with a healthy tree in the same area. Discoloration of the foliage was used as a secondary criterion; the five assessment classes are listed in Table 1 (see over).

In 1984 one third (32.9%) of West Germany's forest was classified as suffering from 'slight' damage and 17.3% as suffering 'moderate' or 'severe' damage¹². This picture did not alter markedly in subsequent surveys, the percentages in the various damage classes remaining almost unchanged (Fig. 1; see over)¹³.

At first, the classification system introduced in Germany defined trees as damaged once the estimated foliar loss exceeded 10%. This threshold was criticized as unrealistic because it ignores the natural variability in crown condition of healthy trees in different regions. The damage classes have thus been changed, and what was initially defined as slight damage (foliar loss of 11–25%) has since 1986 been described as an early warning stage (*Frühwarnstufe*). This redefinition means that less than 20% of the forest area is now classified as damaged, as opposed to just over 50% under the old system (Fig. 1). A second advance in the methods and interpretation of the West German Forest Damage Surveys concerns an underlying assumption. Early surveys implicated air pollutants as the cause of decline by asking for the inclusion only of damage caused by air pollution. As it became clear that there was no such simple cause-effect relationship, that stipulation was dropped. Unhappily, however, the original and premature conclusion has for a long time overshadowed the public view of the causes of new-type damage.

West Germany is no longer the only country where forest surveys have become an annual feature. Since 1984 the Forestry Commission in Britain has been carrying out similar surveys¹⁴, in which ten 'crown

density' classes are used to describe the condition of trees. Transferring these results to the German five-class system, a marked increase in damage to the main conifer species has been apparent over the past four years (Fig. 2). In 1987, 48% of Sitka spruce, 44% of Scots pine and 44% of Norway spruce were in the German categories of moderately or severely damaged (needle loss >25%). Beech and oak, which have only been included in surveys since 1986, also showed extensive crown thinning (66% of oak and 57% of beech in German damage classes 2–4). The Forestry Commission, however, argues that "crown density losses of up to 25% or more may occur quite naturally without the tree having been damaged" (ref. 14) and so is suggesting that damage thresholds are not clear cut. The Commission's analysis of the 1987 survey data showed that

the appearance of trees [that is, crown density] is statistically related to a variety of environmental variables. Some of the explanatory variables examined in the analysis are pollutant levels and there is a general increase in the crown densities of Scots pine, oak and beech with increasing levels of most pollutants. This is entirely consistent with the many experimental studies that have shown that at relatively low levels of atmospheric pollution the growth of plants may be stimulated. The variables used in the study are all highly inter-correlated, placing severe limitations on the extent to which the statistical associations can be interpreted [ref. 15].

Other European countries adopted similar forest survey strategies, and pan-European standardized surveys were initiated under the United Nations Economic Commission for Europe (UNECE) Convention on Long-Range Transboundary Air Pollution. The first report, which is based on the results of the national surveys carried out in 1986, has only recently been published¹⁶. There are considerable differences in the survey methods used and the report does not take the change in the definition of damage (that is, foliar loss in excess of 25% rather than 10%) into account. This compilation of data, however, shows that some defoliation of tree crowns is found in most countries (Fig. 3). The report stresses that no conclusion about the cause of defoliation can be drawn.

A second development is the improved data on the temporal pattern of forest damage. Initial projections of worsening damage that would eventually lead to the destruction of entire forest ecosystems⁴ have not been borne out. Drastically defoliated and dead trees were and are scattered amongst less affected or unaffected trees, and there has been no large-scale deforestation like that seen in areas affected by exposure to high levels of sulphur dioxide in Eastern Europe. The 1987 West German Forest Damage

Table 1 Damage classification used in the annual West German Forest Damage Inventories*

Damage class	Classification	Loss of foliage
0	Healthy	≤10%
1	Early warning stage†	11–25%
2	Moderate damage	26–60%
3	Severe damage	>60%
4	Dead	—

* Extensive yellowing of foliage may lead to reclassification into a worse damage class.

† Original classification was 'slight damage'.

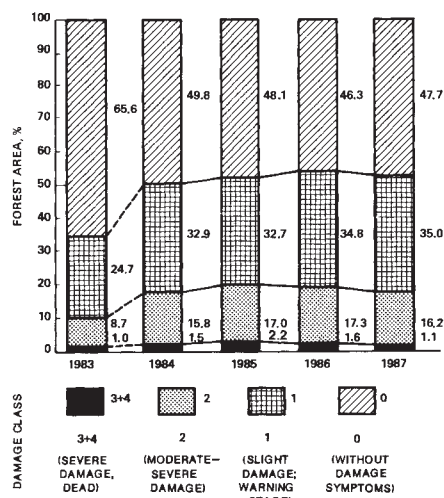


Fig. 1 Results of the West German Forest Damage Surveys 1983–1987. Percentages in the various damage classes have remained almost unchanged since 1984.

Survey concluded that although damage increased in some areas (for example, the Bavarian Alps) and for some species (beech, oak), other areas (the Black Forest, parts of lowland Bavaria) and species (fir, spruce, pine) were showing signs of recovery¹⁵. This conclusion was supported by detailed studies in the Bavarian Forest¹⁷. Such regional differences are concealed by the overall damage figures, which have changed little since 1984 (Fig. 1). Regional improvement shows that damage is reversible if trees are not severely affected. It is therefore high time to stop using the term *Waldsterben* (forest death).

The possible economic effects of forest damage are of course closely linked to projections of the course of that damage. Thus far, fears of severe economic consequences have not been realized. There is no difference in the physical characteristics of timber from damaged and healthy trees¹⁸; and although there has been an increase in the rate of unplanned salvage felling, the excess has been absorbed by the timber market. Prices on the West German market, for example, are largely stable, and there has been no significant decrease since 1982 (refs 19, 20). Of greater concern, however, are the potential long-term effects. Tree-ring studies

show that growth is largely related to climatic factors such as temperature and water availability, which is why there are considerable regional differences in growth patterns. Growth reductions since the 1950s or 1960s have been noted in some areas but there are also reports of marked growth increases in others^{21–23}.

Putting the regional differences to one side, it is important to resolve whether defoliated and healthy trees have significantly different growth patterns and, if so, to what extent growth and defoliation are correlated. The few data available show that, for example, coniferous trees grow more slowly only if they lose more than 30% of their foliage²³. This suggests a threshold rather than a linear relationship, but more information is needed before we can establish a reliable correlation.

All these uncertainties rule out any accurate long-term projections of forest growth. It is therefore difficult to give credence to predictions of forest damage in West Germany for the period 1984–2060(!) which suggest a total monetary loss of as much as 344 billion DM (some \$200 billion), based on loss to forestry, decrease of recreational value and additional costs such as prevention of soil erosion and protection from avalanches²⁴.

Despite the well-documented regional differences in damage symptoms, climate, pollution exposure and so on, the idea of one decline with one cause was at first widely accepted. In 1986, the West German government's Advisory Council on Forest Damage/Air Pollution re-evaluated the evidence and adopted the concept of regional damage types¹. The general acceptance of this principle is the third development in our understanding of new-type forest decline. Five damage types were identified for Norway spruce, one for silver fir and one for beech.

The concept of damage types is a notable advance in thinking, and leads to a more precise description of damage and relevant environmental conditions, thus allowing a better evaluation of possible explanations for forest decline. The simultaneous appearance of various regional damage types within a short period of time over almost the whole of West Germany, suggests that there might be a common synchronizing factor. Any explanation for an individual damage type will thus also have to consider the simultaneous onset of other types of damage.

Some of the five types of spruce damage (Table 2) are well documented, particularly type 1, and a great deal of research has been dedicated to revealing their causes. Five hypotheses have dominated research and are at various stages of evaluation: multiple stress (which relates to damage in general); soil acidification and aluminium toxicity (damage in general); interaction between ozone and acid mist (type 1 spruce damage); magnesium deficiency (type 1

spruce damage); and excess nitrogen deposition (type 5 spruce damage, and a contributory factor to other damage types).

The multiple stress hypothesis proposes that a cupful of stresses — "air pollution with associated atmospheric deposition of nutrient, growth-altering, or toxic substances" (ref. 4) — impairs plant metabolism, making the tree more susceptible to an array of other stress factors (climate, nutrient deficiencies, secondary or contributing pathogens). This hypothesis does not relate to a single tree species or to a specific damage type.

The concept of multiple stress was initially tempting, as it is vague enough to apply to any damage which cannot be explained in any other way. Its very versatility, however, makes it impossible to test and useless to those looking for the causes and mechanisms of forest damage. So here we concentrate on the remaining four, more specific hypotheses.

That invoking soil acidification and aluminium toxicity²⁵ is based on a long-term nutrient-cycling study in the Solling area of Germany. The proposal was that the natural acidification of forest soils is accelerated by the wet and dry deposition of sulphuric and nitric acids from anthropogenic sources (largely combustion of fossil fuels). Increased soil acidity, which leads to the breakdown of the soil buffering system and the release of aluminium ions at levels toxic to plants, would impair the tree's root system and eventually harm the entire tree. This hypothesis, introduced well before the concept of regional damage types, did not address a specific damage type, and conclusions drawn from the Solling study were used to explain damage occurring elsewhere. This extrapolation has been broadly criticized, mainly on the grounds that damage is found in trees growing in a variety of soils, including those that are well buffered against acidification (for example, in the calcareous Alps). There is also evidence that the levels of aluminium found in acid soils are unlikely to cause damage to the roots of the conifer species concerned²⁶. Soil aluminium toxicity is therefore generally not considered to offer a complete explanation for forest damage as a whole or even for a specific damage type. The underlying soil acidification has been suggested as a 'contributing factor' for types 1, 2 and 5 where nutrient deficiencies are involved.

A third hypothesis — that ozone interacts with acid mist, and probably also climatic factors, to induce or exacerbate various nutrient deficiencies^{27, 28} — has been popular in recent years. The idea depends upon increasing ambient ozone concentrations, particularly at medium-high altitudes (type 1 spruce decline), and there is indeed some evidence for a slight overall increase from the 1950s through to

the 1970s. But it is not clear that this trend continued in the 1980s. There has been a great deal of experimental work on this hypothesis, but these field and laboratory studies have so far failed to provide supporting evidence for it. The damage symptoms produced by ozone/acid mist in the laboratory (with or without climatic stress) are not the same as those seen in the field. Furthermore, the ozone generators used in most studies have been shown to produce significant quantities of nitric acid¹², and previous results on cation leaching from needles will have to be reassessed. On the strength of the evidence, we would say that where nutrient deficiencies are involved (for example spruce damage types 1–4) the role of ozone has been overestimated.

The fourth hypothesis also focuses on one specific type of damage to Norway spruce (type 1). Various reports have stressed the role of magnesium deficiency for this 'high-altitude spruce damage' (refs 26, 33). Mass balances of magnesium for German sites show that magnesium nutrition was probably always marginal. There is limited evidence that, with soil acidification, exchangeable soil

magnesium has declined in some forest soils because of tree uptake and leaching by acid deposition. The sudden occurrence of symptoms over a wide area is attributed to a series of dry years which reduced magnesium mineralization and uptake. If insufficient magnesium nutrition is indeed at the heart of the problem, then type 1 spruce damage could hardly be classified as new-type damage, although it is clear that these damage symptoms now occur over a larger area than ever before. Magnesium depletion caused by tree harvesting and leaching of soils by sulphur and nitrogen deposition, as well as the reduced uptake of magnesium caused by drought, may offer the explanation.

Type 1 damage is, however, restricted to inland continental Europe, and it is unlikely that the magnesium deficiency and related forest damage will occur in countries along the seaboard of Northern Europe. The rate of atmospheric magnesium deposition from maritime sources, and consequently the magnesium status of upland forests in these countries (for example Britain and Norway) is much greater than the loss of magnesium caused by harvesting or leaching²⁶.

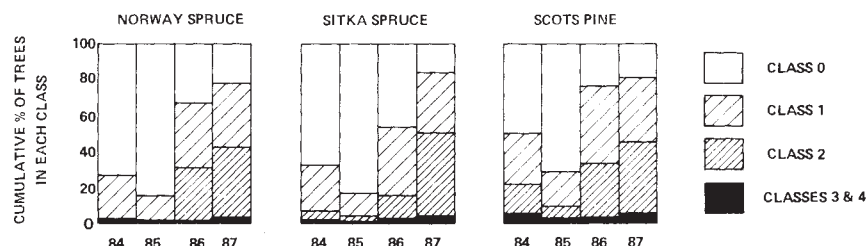


Fig. 2 Results of the British Forestry Commission's surveys 1984–1987, transferred to the West German five-class system, showing a marked increase in damage to the main conifer species.

Table 2 Decline types for Norway spruce (*Picea abies* L. Karst)¹

Type	Official description	Main symptoms	Geographical occurrence
1	Needle yellowing at higher elevations	Chlorosis and subsequent loss of older needles, associated with foliar magnesium deficiency	Higher altitudes of medium-range mountains in central and southern Germany (for example, Black Forest, Bavarian Forest)
2	Thinning of tree crowns at medium-high altitudes	Loss of needles, but little chlorosis	Medium-high altitudes (approximately 400–600 metres), particularly in central and northern Germany (for example, Hils, Odenwald)
3	Needle reddening of older stands in southern Germany	Yellowing, then reddening of older needles, followed by needle shedding	Mainly in the foothills of the Bavarian Alps
4	Chlorosis of needles in the Bavarian Alps	Chlorosis and subsequent loss of needles, starting with the youngest needles. Associated with K and/or Mn deficiency	Calcareous Alps, above approximately 1,000 metres
5	Thinning of tree crowns in northern coastal areas	Reduced growth, loss of needles	Northern German Plain

The fifth and final hypothesis considered here — damage caused by excess nitrogen deposition¹¹ — has gained ground in the past few years. Although nitrogen availability is often a limiting factor in plant growth, deposition in excess of requirements may increase susceptibility to climatic and biotic stresses, and accelerate soil leaching^{35, 36}. These effects have been observed in pine stands in the Netherlands³⁷, where nitrogen deposition can exceed 50 kg N ha⁻¹ yr⁻¹, mainly in the form of ammonia released from animal waste. Local ammonium deposition may thus be an important factor in type 5 spruce damage (see Table 2). It is less obvious that nitrogen deposition has a similar role in other types of damage at higher altitudes in central and southern Germany, where deposition rates are much lower (less than 25 kg N ha⁻¹ yr⁻¹) and the ratio between nitrate and ammonia/ammonium may be different. Increased growth of spruce may have increased the annual magnesium requirement. This growth increase, observed in the Black Forest since 1965 (ref. 23), may have been linked to favourable climatic conditions and possibly to increased nitrogen deposition.

So, what have been the recent advances in our perception of the new-type forest decline?

- A more realistic definition of damage takes the natural variability of tree condition into account and sets a generally accepted threshold for damage (loss of foliage in excess of 25%).

- Initial projections of continuously increasing damage have not been borne out. There is, on the contrary, regional recovery.

- It is now generally accepted that the new-type forest damage is not a single phenomenon, but a series of regional damage types, possibly triggered by a common synchronizing factor.

- The causes of the individual damage types are partly resolved. Soil magnesium deficiency appears to be the key factor in type 1 spruce damage, exacerbated by tree harvesting, acid deposition and drought. Excess nitrogen deposition could, at least in principle, play a part, whereas the involvement of ozone now looks less certain. Needle-reddening of spruce (type 3 damage) is primarily caused by needle-cast fungi. Local ammonium deposition is probably one cause of type 5 damage in northern Germany. The causes of type 2 damage (crown thinning at medium-high altitudes) have not been resolved. The same applies to spruce damage in the calcareous Alps (type 4), although foliar potassium and manganese deficiency are involved.

More generally, there has been a welcome trend towards a less emotional and more reasoned approach to forest decline. In West Germany, interest in the subject is fading in the wake of other

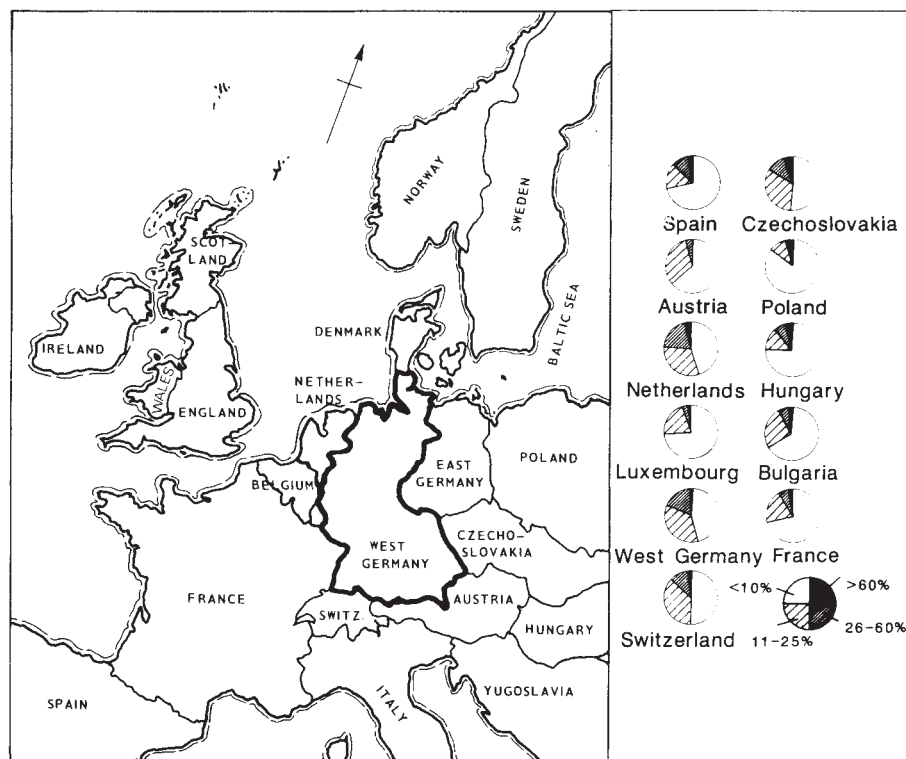


Fig. 3 Defoliation of European forest — results of the 1986 UNECE Forest Health Survey, transferred to the West German five-class system, indicating that defoliation is widespread over Central Europe. Percentages are of foliage loss observed, and correspond to the damage classes in Table 1. Results of the British survey of 1986 (see Fig. 2) are not included because of differences in methodology.

environmental upsets such as the consequences of Chernobyl and pollution of the North Sea, and because it has become clear that there is to be no nationwide *Waldsterben*. The situation in Britain and North America is somewhat different, as the condition of forests has only comparatively recently become an issue. But that is in large part an advantage — those working on forest decline can draw upon experience and scientific results from elsewhere, thereby avoiding the early misconceptions and putting the public debate on a firm factual footing. □

1. Forschungsbeirat. *Waldschäden/Luftverunreinigungen* (Report 2, Karlsruhe 1986).
2. Blank, L. *Nature* **314**, 311–314 (1985).
3. McLaughlin, S. B. J. *Air Pollut. Control. Ass.* **35**, 512–534 (1985).
4. Schütt, P. & Cowling, E. B. *Pl. Dis.* **69**, 548–558 (1985).
5. *Acidic Precipitation* (Kommission Reinhaltung der Luft, VDI, Düsseldorf, 1983; English edn, 1987).
6. Innes, J. L. *Forestry Commission Bulletin* 70 (HMSO, London, 1987).
7. Wachter, A. J. *Pl. Dis. Prot.* **85**, 361–381 (1978).
8. Wislizenus, H. (ed.) *Waldsterben im 19. Jahrhundert* (VDI, Düsseldorf, 1985).
9. Bauer, F. (ed.), *Diagnose und Klassifizierung der Neuartigen Waldschäden in Die Sache mit dem Wald* (Bayerischer Landwirtschaftsverlag, Munich, 1985).
10. Wagner, F. *Forstw. Cbl.* **100**, 148–160 (1981).
11. Moriondo, F. & Covassi, F. *Forstw. Cbl.* **100**, 168–170 (1981).
12. *Forest Damage Survey*, (Federal Ministry for Food, Agriculture and Forestry, Bonn, 1985).
13. *Forest Damage Survey*, (Federal Ministry for Food, Agriculture and Forestry, Bonn, 1987).
14. Innes, J. L. & Boswell, R. C. *Forestry Commission Bulletin* 74 (HMSO, London, 1987).
15. Innes, J. L. & Boswell, R. C. *Forestry Commission Bulletin* 79 (HMSO, London, 1988).
16. *Forest Damage and Air Pollution* (Report of the 1986 forest damage survey in Europe, UNEP and UNECE, Global Environment Monitoring System, Geneva, 1987).

17. Kandler, O., Miller, W. & Ostner, R. *Allg. Forstz.* **27/28**, 715–723 (1987).
18. Ammer, U. & Kenneweg, H. *Schadsymptome, Inventur [Fernerkundung], Holzzuwachs, Holzqualität*, 124–144 (Proc. Statusseminar Ursachenforschung Waldschäden, KfA Jülich, 1987).
19. Wilkening, W. *Fakten und Tendenzen am Internationalen Holzmarkt* 119–212 (KIK-PEF 18, KfA Karlsruhe, 1986).
20. Hein, W. & Bitter, W. G. *Allg. Forstz.* **7**, 132 (1988).
21. Schweingruber, F. H. *Proc. Symp. Klima und Witterung in Zusammenhang mit den neuartigen Waldschäden*, October 1986, 165–174 (Munich, 1987).
22. Eckstein, D. & Bauch, J. *Proc. Symp. Klima und Witterung in Zusammenhang mit den neuartigen Waldschäden*, October 1986, 156–164 (Munich, 1987).
23. Kenk, G. & Fischer, H. *Env. Pollut.* **54**, 199–218 (1988).
24. Schulz, W. *The Economic Costs of Air Pollution* (Lecture, Goethe-Institute, London, 1987).
25. Ulrich, B. in *Effects of Air Pollutants in Forest Ecosystems* (Ulrich, B. & Pankrath, J. eds) 1–32 (Reidel, 1983).
26. Roberts, T. M., Skeffington, R. A. & Blank, L. W. *Forestry (UK)* (in the press).
27. Prinz, B., Krause, G. H. M. & Stratmann, H. *LIS Report* 28 (Landesanstalt für Immissionsschutz, Essen, 1982).
28. Bosch, C., Pfankuch, E., Baum, U. & Rehfuess, K. E. *Forstw. Cbl.* **102**, 167–181 (1983).
29. Krause, G. H. M., Jung, K. D. & Prinz, B. *VDI Berichte* **560**, 627–656 (1985).
30. Skeffington, R. A. & Roberts, T. M. *Oecologia* **65**, 201–206 (1985).
31. Rehfuess, K. E. & Ziegler, H. *Forstw. Cbl.* **105**, 267–271 (1986).
32. Brown, K. A. & Roberts, T. M. *Env. Pollut.* **55**, 55–73 (1988).
33. Zöttl, H. W. & Hüttl, R. F. *Wat. Air Soil Pollut.* **31**, 450–462 (1986).
34. Skeffington, R. A. & Wilson, E. J. *Env. Pollut.* **54**, 159–184 (1988).
35. Agren, G. I. in *Ecological Effects of Acid Deposition* 233–244 (National Swedish Environment Protection Board, Stockholm, 1983).
36. Nihlgård, B. *Ambio* **14**, 2–8 (1985).
37. Roelofs, J. G. M., Kempers, A. J., Hudijk, A. L. F. M. & Jansen, J. *Pl. Soil* **84**, 45–56 (1985).

L. W. Blank, T. M. Roberts and R. A. Skeffington are at the Central Electricity Research Laboratories, Kelvin Avenue, Leatherhead, Surrey KT22 7SE, UK.

Modelling the evolution of globular star clusters

Piet Hut*, Junichiro Makino† & Steve McMillan‡

* Institute for Advanced Study, Princeton, New Jersey 08540, USA

† Department of Earth Science and Astronomy, College of Arts and Sciences, Tokyo University, Komaba, Meguroku, Tokyo 153, Japan

‡ Department of Physics and Atmospheric Science, Drexel University, Philadelphia, Philadelphia 19104, USA

Globular star clusters provide a unique laboratory. In astronomy they present an opportunity to study dense stellar systems that are more accessible than galactic nuclei. In statistical mechanics, concepts of negative heat capacity and resulting gravothermal instabilities challenge the present framework of statistical descriptions of dynamical systems. In computer science, modelling their evolution poses an extreme challenge to the hardware and software capabilities of the next generation of parallel computers, and provides an ideal test case for teraflop machines.

THE structure and evolution of the two hundred or so globular star clusters that circle around our Galaxy presents an intriguing problem. As each cluster has $\sim 10^5$ – 10^6 member stars, their dynamics are too complex for detailed star-by-star simulation on present-day computers, yet their long-term evolution is believed to be critically influenced by episodes whose proper description mandates just such an approach¹. Over the past two decades, a wide variety of mathematical modelling techniques has been developed that yields a wealth of physical insight into cluster dynamics². But these methods are currently reaching the limits of their applicability, and further advances will soon require detailed modelling with a qualitatively different degree of realism. What is needed now is nothing short of a star-by-star modelling of globular clusters. Formidable though such simulations may be, they will provide significant benefits for astrophysicists, statistical physicists and computer scientists alike.

For astrophysicists, globular clusters provide an opportunity to observe dense stellar systems in action, without the obscuring effects of gas and dust which impede similar observations in galactic nuclei³. Ground-based optical measurements of increasing quality, coupled with anticipated high-resolution observations with the Hubble Space Telescope, will constrain permissible theoretical models considerably¹. In addition, X-ray binaries and millisecond pulsars provide us with welcome observational handles on the unusual dynamical environment in globulars, as these systems are probably formed through physical encounters between stars⁴.

For statistical physicists, a globular cluster, in its simplest form, can be modelled as a self-gravitating collection of mass points, interacting simply through inverse-square gravitational attraction. Although simple in principle, this is a many-body problem that provides an extreme challenge to statistical mechanics. The qualitative evolutionary behaviour of such a system is now more or less understood. Two-body encounters slowly redistribute energy, transporting heat from the centre of the cluster to the outskirts. As a consequence of the loss of kinetic energy, the inner regions contract. But a slow contraction along near-equilibrium states of a self-gravitating system necessarily heats the contracting part (as a consequence of the virial theorem), accelerating the heat flow from the central core into the cooler, expanding outer halo. This mechanism produces an infinite central density in a finite time, at least in the two-body encounter approximation. This breakdown of the two-body formalism signals the importance of three-body and higher order interactions during the late stages of core contraction. In fact, the formation of binaries will ultimately reverse the collapse and cause core re-expansion, a process that has begun to be explored in detail only during the past few years¹.

This process of core collapse is driven by the so-called gravothermal instability, which may be interpreted as a manifestation of the negative heat capacity of a gravitating system.

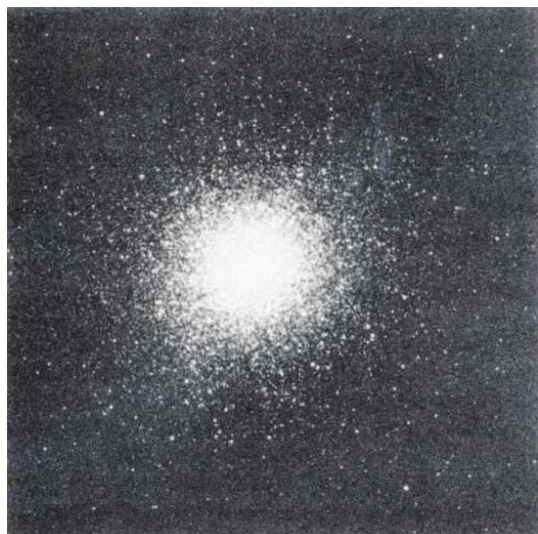
A simple illustration of this effect is the paradox of a satellite encountering the Earth's atmosphere, where frictional braking leads to an increase in speed. Strictly speaking, we cannot apply to gravity the usual terminology of thermodynamics, including concepts such as heat capacity, because long-range forces are explicitly excluded from the standard framework of statistical physics. Thus, the quasi-static evolution of a globular cluster still awaits a proper formalism, and the extension of current techniques to problems of this nature remains one of the central challenges to statistical physics.

Even the simple mass-point approximation offers a bewildering diversity of behaviour in addition to the simple picture sketched above. During the process of core expansion, the original gravothermal instability sets in again, causing gravothermal oscillations of the core. These oscillations were discovered by Bettwieser and Sugimoto⁵, who illustrated their chaotic nature. It was both unexpected and fascinating to find this type of behaviour in so basically simple a system as a set of ($\sim 10^6$) mass points whose interactions are described simply by a mutual inverse-distance potential.

Lastly, modelling globular cluster evolution poses an extreme challenge for computer scientists. Currently, computer hardware is not fast enough to follow the complete evolutionary history of self-gravitating systems containing more than $\sim 10^4$ particles. But with a (relatively) modest advance in computer speed (of a few orders of magnitude⁶), it will become possible to perform globular cluster simulations by explicit integration of the orbits of individual stars, instead of having to resort to the approximate methods now commonly employed. Such simulations involve the modelling of long-range forces which act simultaneously over length scales spanning 15 orders of magnitude and time scales spanning 19 orders of magnitude (J.M. and P.H., manuscript in preparation). This poses extreme requirements, not only on the hardware, but also on the software used to model these systems. The gravitational many-body problem thus proves to be an ideal testing ground for developing algorithms and hardware for handling these extreme requirements of dynamic range in space and time.

Computer simulations of globular clusters

In many respects, models of globular cluster evolution have reached a point comparable to the state of stellar evolution theory in the 1950s, when nuclear reaction rates describing the physics of energy generation became available and the computers necessary for the construction of detailed models were just being developed. The analogous two- and three-body gravitational reactions between stars in globular clusters have recently been studied in detail. These reactions turn on after the initial core collapse and heat the central regions of the cluster both directly, through the effects of energetic reaction products, and indirectly, through mass loss. They power the 'main-sequence'



Dynamical evolution of globular clusters

Globular clusters are systems of typically hundreds of thousands of stars which are nearly in dynamical equilibrium. Simple as these systems appear, their dynamical evolution poses several unsolved problems. On timescales of billions of years, heat generation through star-star interactions leads to a flow of energy from the centre to the outskirts, causing a contraction of the innermost regions of the cluster, a phenomenon known as core collapse.

When the core reaches a sufficiently high density, a variety of energy-generating mechanisms can cause it to re-expand. This re-expansion is generally unstable, leading to alternating periods of core contraction and core expansion with a seemingly chaotic long-term character.

Illustrated here is the globular cluster M13 (NGC6205), which was observed by Halley in 1714. It is estimated to contain about 500,000 stars.

phase of a globular cluster and drive a slow but steady loss of stars by evaporation. Computationally, the present development of new super- and parallel computers promises, within a few years, speeds many orders of magnitude greater than those available today, which should enable explicit study of more complex many-body behaviour.

The progress in our understanding of globular cluster dynamics during the last three decades has been considerable. We now have a consistent standard picture of pre-collapse evolution, initiated during the 1960s and developed in detail during the 1970s. Some of the main ingredients of post-collapse evolution have emerged during the 1980s. Spitzer² has recently provided an excellent introduction to the (mostly pre-core-collapse) dynamical evolution of globular clusters. But major questions about the post-collapse phase remain unanswered. Our insight into the further stages of evolution, during and after core bounce, is much less complete. Fundamental questions, such as whether gravothermal oscillations will occur in realistic cluster models, are largely unsettled. We are still far from the point where we can construct models that can be compared directly with observations.

A severe limitation to our current understanding of the later stages of globular cluster evolution lies in the nature of the numerical simulations that have been carried out. They have been largely based on a description in terms of statistical diffusion, such as a slow drift of particle orbits in energy space, described by a Fokker-Planck approximation. Although these methods have been fruitful for furthering our understanding of pre-collapse evolution, they have severe shortcomings when applied to the later phases. One limitation lies in their statistical nature, which is of questionable validity after core collapse, when the core may contain less than a hundred particles and typically only one or zero close binaries. A second limitation for realistic applications lies in the limited number of physical parameters that can be modelled. For example, providing statistical bins for a simultaneous collection of mass, energy and angular momentum choices will result in many bins containing only a fraction of one star, obviously jeopardizing the statistical assumptions of a Fokker-Planck algorithm. Current models for star cluster evolution necessarily neglect many important astrophysical effects that are automatically included in a star-by-star calculation. Examples are the intermittency of energy generation after core collapse, the possible occurrence of non-radial oscillations in the central regions, fluctuations in heat transport arising from the small number of stars in and near the core, and the combined effects of mass segregation and anisotropic velocity distributions.

Even in the simplest case of the pure N -body problem, the long-term evolution has been followed numerically only for $N \leq 3,000$. We believe that future cluster models must extend this regime to $N \sim 10^5$, the size of an average globular cluster; we discuss below the advances in hardware and software that may be necessary to achieve this. A convenient round number for the smallest number of stars in a realistic simulation of globular cluster evolution might then be $N = 2^{16} = 65,536$ (the power of two is chosen with the possibility in mind of running such a simulation on a parallel computer based on a hypercube architecture, where each star could be mapped directly onto a processor). Independent motivation for an increase to this particular size is provided by the work of Goodman⁷, who finds that the behaviour of a system of mass points is qualitatively different below and above certain critical values of N in the range $(0.7-4) \times 10^4$.

Hybrid techniques

The question we must then tackle is how to address the many limitations just described. We would like to model at least the central region of a cluster, out to a few core radii, say, on a direct star-by-star basis. This would allow us to increase the realism of the simulations, by taking into account the combined effects of star evolution, individual mass loss, mergers, tidal encounters and the presence of white dwarfs and neutron stars, which react very differently in physical collisions. In addition, any type of binary can be allowed to form and evolve without pre-imposed constraints or assumptions. An important issue is whether there are any shortcuts—can we achieve a realistic modelling of the core without paying the price of following all 10^5 stars of a globular cluster individually?

The first realization of such a hybrid approach, combining high resolution in the centre, down to the level of individual stars, with a statistical description of all but the inner few hundred stars, was described by McMillan and Lightman⁸ and McMillan^{9,10}. Their method demonstrated the basic feasibility of mixing N -body and statistical techniques, and showed how few-body effects are intimately involved in the termination of core collapse for even a large- N system. In their simulation, the post-collapse core underwent a series of oscillations, whose energetics could be completely explained as the direct effects of binary heating and ejection. The oscillations did not seem to be gravothermal in nature, a conclusion strengthened by their small amplitude and relatively short timescales.

The apparent absence of gravothermal oscillations is contrary to the predictions of Fokker-Planck and gas sphere simulations, as the effective number of stars in the simulation was formally

infinite. A possible explanation is simply that the hybrid integration covered just the early portion of a long-timescale gravothermal event and that the expected long-term behaviour was masked by short-term noise. But the picture is complicated by the presence of order-unity fluctuations in both the stellar density and the velocity dispersion in the core. Both of these quantities are necessarily taken to be well represented by the mean-field approximation in the Fokker-Planck formalism, and the possibility that large stochastic variations in them might somehow suppress post-collapse gravothermal oscillations cannot yet be ruled out.

The economic facts of life

The main limitation in running a hybrid (or any other N -body) simulation is computational expense. The longest hybrid run made so far covered a post-collapse period of only about 0.005 initial half-mass relaxation times. By comparison, the period of the oscillations found by Cohn *et al.*¹¹ is typically several hundred times longer. Clearly, more extensive N -body simulations are required if the post-collapse evolution is to be covered adequately.

We can get a rough idea of how much computer time can be saved using a hybrid approach by examining the way in which computational effort is distributed through an N -body calculation. Figure 1 indicates the fraction of computer time expended on the stars comprising the innermost fraction q of the total mass. In applications to globular clusters in the post-core collapse phase, the inner few per cent of the particles will have a density profile in between a strictly isothermal one (left behind during a re-expansion phase), and one with a density exponent $\alpha \approx 2.23$ (left behind during a contraction phase). The midpoint of the computational effort shifts during core oscillations between inner mass fractions of 0.4–15% (Fig. 1). During this period the core radius also changes, in such a way that the midpoint of computation remains at a few core radii (J.M. and P.H., manuscript in preparation). The conclusion is that the total amount of computation will always be divided in roughly comparable amounts between the core, a region just outside the core (containing a comparable number of stars) and the rest of the system (containing most of the mass).

Two implications follow immediately. First, during periods of extreme core contraction, the computational cost per unit of physical time increases by a factor of less than five, the maximum value of which is given by the ratio of the slopes of the two lines in Fig. 1 for $q \rightarrow 1$. This implies that we can determine the total cost of the calculation of a globular cluster evolution, to order of magnitude, without worrying about the details of the core structure. Second, the gain expected from a hybrid algorithm, which makes very large relative savings in the region beyond a few core radii, is unlikely to make a significant absolute saving, as about half of the total computational effort is spent in the central region, which is modelled star-by-star. Compared with the total speed-up needed, even an order of magnitude improvement in the force calculation cost goes only a small way towards reducing the total expense of a complete run (see below).

We now address these issues in relation to the next generation of globular cluster simulations by discussing different methods of approximation, algorithms to implement them and the requirements for the hardware needed to run the models.

The next generation of simulations

To understand what it takes to model a globular cluster on a star-by-star basis, and before exploring possible shortcuts, we must discuss the different categories of N -body simulations relevant to star-cluster simulation.

The simplest way of modelling a system of N particles is by computing all N^2 pairwise interactions at regular short intervals. This lockstep method of shared time steps is terribly inefficient, even if the step size is allowed to change with time, especially

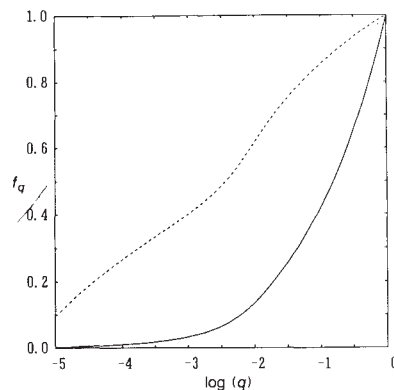


Fig. 1 Cumulative fraction of computer time, f_q , spent on the particles that constitute the innermost fraction q of the whole cluster, for cluster density profiles of the form $\rho \propto r^{-\alpha}$, with r the distance to the centre. This estimate is based on the use of the Ahmad-Cohen scheme, a two-timescale method, with constant neighbour number (compare ref. 14). The solid line is for a singular isothermal particle distribution ($\alpha = 2$), whereas the dashed line indicates a steeper particle density ($\alpha = 2.25$) close to that occurring during the later phases of core collapse². The total number of particles is $N = 100,000$.

when one or more binaries are present, because they bring down the shared time-step size by extra orders of magnitude. Performed in this way, a star cluster calculation with 65,000 stars would take a time equivalent to the age of the Universe on a gigaflop supercomputer.

A much more efficient scheme was introduced by Aarseth^{12,13}. Here, each particle is represented by a polynomial fit to an orbit segment, making individual time steps possible but still allowing particles to interrogate each other at random intermediate intervals. The first detailed analysis of the computational complexity of such a scheme¹⁴ provided explicit expressions for the scaling of computational cost with particle number and steepness of density distribution. An extension of these results, which includes an analysis of the additional cost required by binary interactions (J.M. and P.H., manuscript in preparation), leads to the conclusion that a simulation of a 65,000-star cluster requires $\sim 10^{19}$ floating-point operations, that is, about 100 days on a teraflop or a few centuries by gigaflop (see below).

The next significant algorithmic improvement was the introduction of a two-timescale method¹⁵, in which each particle is assigned a neighbour sphere, dividing the other particles into near neighbours and distant friends: neighbour forces are evaluated much more often than those of far friends. This trick of concentrating computation where it is most needed results in a substantial saving in computer time. Using Makino and Hut's analysis¹⁴, we estimate the saving for 65,000 particles to be about a factor of ten (Fig. 2).

What further algorithmic improvements can be expected? There are three general categories of possible methods. The first is the most straightforward extension of the Ahmad-Cohen scheme¹⁵, whereby for each particle, all other particles are partitioned amongst three or more neighbourhoods, for example, in the form of concentric shells, and different time-step sizes are assigned to the interaction with members of different neighbourhoods.

The second category extends the flexibility of regrouping interactions in time to a regrouping of interactions in space and time simultaneously: interactions with distant particles are not evaluated individually but instead in the form of a multipole expansion of whole groups of particles at the same time. The book-keeping structure needed to do this typically takes on the form of a tree hierarchy and we will refer to this wider class as tree methods. They include generalized neighbour schemes as special cases.

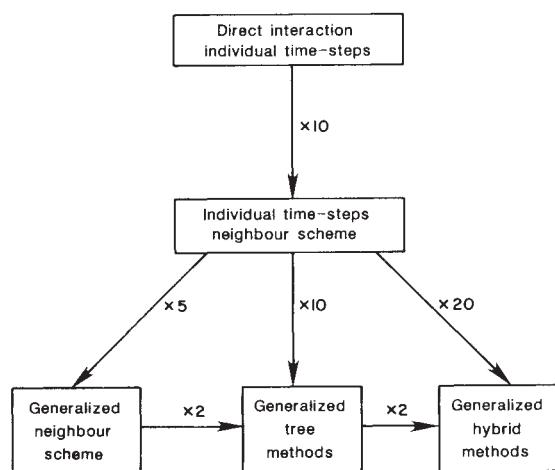


Fig. 2 Relative speed-up in computer time needed to reach the same accuracy, for variants of direct N -body simulations, for a calculation with 50,000–100,000 particles. Computational speed increases from top to bottom and from left to right. We estimate these speed-up numbers to be accurate to within a factor of two.

The third category goes beyond regrouping only of forces in space and time, by lumping together even the basic information about whole groups of stars, treating them as single composite entities rather than as individual particles. If we were to treat stars in the core in this way, we would return to a purely statistical description in some type of generalized Fokker-Planck form. Stars beyond the inner few core radii probably can be treated statistically without significant loss of accuracy. These stars comprise 95–97% of the cluster, at least in the post-core collapse phase, as gravothermal core oscillations involve only the inner 1–2% of the stars (refs 7, 11 and D. C. Heggie and N. Ramamani, preprint).

At present, to the best of our knowledge, none of these three improvements has been implemented in a method capable of simulating globular cluster evolution. The first, using a higher-order neighbour scheme, is straightforward to implement in principle, although the book-keeping will become rather tedious. The second, using a tree method, will require extensive experimentation to ensure that no spurious relaxation is introduced. The third, using a hybrid approach, would entail the addition of an inner-neighbour scheme to the existing version; although this is quite easy to accomplish, the artificially sharp interface between the N -body and statistical regimes will continue to raise questions of consistency in any very long run.

In Fig. 2 we estimate the relative speed-up gained in switching between the present standard neighbour scheme and each of these three improvements. These numbers apply for 65,000 particles, with an estimated uncertainty in each speed-up value of a factor of about two. They were obtained by a combination of the following three arguments.

First, extending Makino and Hut's analysis¹⁴ to additional neighbour spheres is straightforward, although cumbersome. Without reproducing here the details of the derivation, we can offer the following heuristic argument for the factor of five speed-up gained from introducing higher levels of neighbour hierarchy (Fig. 2). Break-even between zero and one neighbour level, at the relatively high accuracy required for globular cluster calculations, occurs around $N \sim 2^7$. Break-even between one and two neighbour levels should occur at values of N somewhat below the point where a neighbour sphere contains 2^7 particles, which occurs around $N \sim 2^{13}$ (ref. 14). For $N \sim 2^{16}$, the speed-up gained from the second level of neighbours is expected to be comparable to that achieved at $N \sim 2^{10}$ from the introduction of the first, that is, about a factor of five. Although a more detailed analysis is needed to justify these numbers, these simple

considerations show that we are unlikely to gain more than one order of magnitude in speed-up.

Second, the factor twenty presented in Fig. 2 for the speed-up that we expect would be introduced by generalized hybrid methods is based on the requirement that at least the region immediately surrounding the core—say, within three core radii—be treated exactly. Whereas at the point of collapse the core contains only a few dozen stars, its peak size during the cycle of gravothermal oscillations is perhaps 1–2% of the total cluster mass. Clearly, it is not consistent to treat only a portion of a roughly homogeneous core with the N -body integrator, so $\sim 5\%$ of the particles must be modelled directly. As most of the computer-time requirements are associated with these particles, the speed-up over a non-hybrid integration is given simply by the reduction in the cost of the force-evaluation loop, that is, a factor of twenty. It is important to note that the speed-up in a hybrid integration is realized only if the inner stars feel no force arising from the outer regions, that is, only for a strictly spherically symmetric cluster. Any attempt to accord to the outer regions some departure from sphericity, by including quadrupole moments, for example, is likely to significantly impair the performance of the method.

Third, the maximum speed-up that can be obtained with a general tree method, whether based on a Lagrangian approach^{16–18} or an Eulerian one^{19–21}, can be estimated as follows. When we allow individual time steps within a tree-based method, we can achieve a speed-up at least as large as can be achieved by a multiple-neighbour scheme, for the latter uses optimization only in the time-grouping of interparticle forces, whereas tree methods give an additional grouping of forces in space. On the other hand, tree methods will always be slower than Fokker-Planck methods in the outer regime of a hybrid method. Thus the speed of the most efficient tree method will always lie between the speeds of the most efficient multiple-neighbour scheme and hybrid method. This is indicated in Fig. 2, within the relative uncertainty of a factor of two, simply by placing the tree methods at the geometric mean speed of the other two types of method.

At this point, there is no reason to carry the analysis to higher accuracy, because additional uncertainties, dependent on the particular hardware used, quickly become more important. Note that the analysis is valid strictly only for scalar computers, such as workstations and mainframes. Calculations for globular clusters containing 2^{16} particles can be carried out only on supercomputers, with a mix of vectorization and/or parallelization being unavoidable. The resultant increase in programming complexity, coupled with a significantly better performance for algorithms with a simpler structure, will certainly lower the numbers for the relative speed-up values presented in Fig. 2. Depending on the hardware available, it is possible that none of the three types of improvement may give any appreciable speed-up, in which case the traditional neighbour scheme might still offer the best strategy. It may even be that the most efficient method will be the even simpler individual time-step approach, in which no neighbour scheme is used at all.

Of the three approaches, the most practical seems to be the implementation of some sort of tree method. Such a method would cut through the layers of book-keeping complexity that seem to be inevitable with the generalized Ahmad-Cohen scheme but avoids the inflexibility and potential inconsistency inherent in an extended hybrid method. In some respects, the tree-based approach is the natural generalization of existing Eulerian hybrid schemes.

To put the speed comparisons (Fig. 2) in perspective with respect to the presently used Fokker-Planck and conducting-gas-sphere approaches, Fig. 3 shows the enormous gap in computer speed that must be bridged before applying any of the methods described above. We use empirical values for the number of floating-point operations required for state-of-the-art simulations with gas methods and Fokker-Planck algorithms. For the

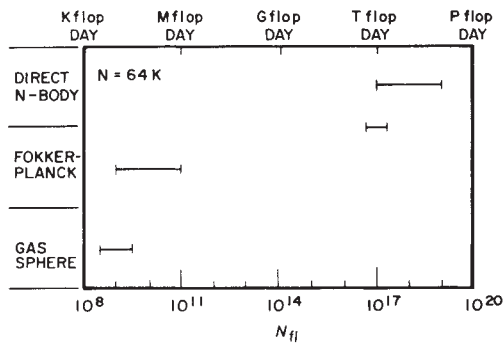


Fig. 3 Computer time required to simulate the evolution of a star cluster through core collapse and a comparable time into the post-collapse phase. Three types of models are indicated: direct N -body simulations, Fokker-Planck calculations and conducting-gas-sphere models. The fourth line, in between the first two, indicates hybrid models which incorporate both direct and Fokker-Planck elements. The number of floating-point calculations is given along the abscissa above and is translated into products of speed and time below. These numbers are our best estimates for a cluster containing $2^{16} \sim 65,000$ particles and are accurate to within an order of magnitude. Increasing N by a factor of two will increase the estimates for direct N -body calculations by an extra order of magnitude.

speed needed for direct calculations and hybrid methods, we use the results from Fig. 2, combined with the fact that the simplest individual-time-step method of direct N -body simulation requires $\sim 10^{19}$ floating-point operations (J.M. and P.H., manuscript in preparation), as shown by the following simple argument. We will use the standard units described by Hénon²² and Heggie & Mathieu²³, in which the total mass of the cluster is unity, the total energy is $-1/4$ and $G=1$. In these units, the half-mass radius and the velocity dispersion are of order unity, independent of N , and the same therefore also holds for the crossing time.

We start with the empirical result that running a system of 1,000 particles to core collapse using Aarseth's NBODY5 code¹³ takes a few hours on a Cyber 205. On this vector machine, the maximum speed is limited to a few tens of Mflops because of the extensive use of indirect addressing. A typical calculation extending well into the post-collapse phase will therefore require $\sim 10^{12}$ double-precision floating-point calculations using an Ahmad-Cohen neighbour scheme. Without the neighbour scheme, the number of calculations would be greater by a factor of about five¹³. Thus, about 5×10^{12} floating-point calculations are required to simulate a star-cluster evolution through core collapse and substantially beyond, using only individual time steps. Increasing the value of N from 1,000 to 65,000 for the individual time-step method, introduces three additional factors that increase the computer time required. First, the number of interactions scales as N^2 . Second, the number of steps required per crossing time increases at least as fast as the inverse interparticle distance, $\propto N^{1/3}$. In fact, it has to scale slightly faster, as we require increasingly accurate computations per step for

higher N , to arrive at a fixed overall accuracy at the end of a run. We assign this extra cost a factor $\propto N^{1/6}$. Third, the number of crossing times per relaxation time scales as N (neglecting a factor logarithmic in N). Combining all these factors, we find a scaling in computer time $\propto N^{3.5}$. The increase from 1K to 64K particles will lead, therefore, to an estimated number of floating-point operations of $5 \times 10^{12} \times 64^{3.5} \approx 10^{19}$.

Simulation of globular cluster evolution

These arguments lead to three conclusions: (1) calculations that model a globular cluster with 10^5 stars on a star-by-star basis, using currently available methods, are necessarily more computer-intensive than current Fokker-Planck methods by a large factor, of at least 10^7 ; (2) hybrid codes, which save computer time by introducing a Fokker-Planck treatment for the main part of the cluster (outside a few core radii) are still about 10^6 times more computer-intensive than current pure Fokker-Planck methods; (3) multiple-neighbour schemes or hierarchical tree algorithms will require an amount of computer time comparable to hybrid algorithms, but have the advantage of modelling the entire cluster on a star-by-star basis. Significant improvement in globular cluster simulations, over and above Fokker-Planck models, thus requires a jump from megaflop-days to gigaflop-years or, equivalently, teraflop-days.

During the next few years, gigaflop speeds will not yet be generally available for sustained periods. Simulations of smaller star clusters containing $\sim 10^3$ – 10^4 particles should, however, be performed in the near future to develop the software necessary to conduct, at a later date, more realistic experiments with 2^{16} stars. The implementation of analysis programs for binary formation and evolution, integrated to some extent with the main-orbit integration program, in particular will require a significant investment both in human time and in computer time to test the concepts on intermediate-sized systems.

Exactly when teraflop speeds will become available is not yet clear but it could be as soon as a few years from now⁶. The globular cluster problem would be an ideal test case to run on such a machine, for the time requirements are nontrivial but not extreme, as one simulation would require about a day of computation; and because the structure of the algorithm used has a significant complexity, which differs in nature from most supercomputer algorithms because of the long-range character of gravitational interactions.

Another interesting possibility would be to build special-purpose hardware dedicated to running simulations of globular cluster evolution—consider, for instance, the digital orrery described by Applegate *et al.*²⁴. This is already feasible technologically and will soon become feasible economically. The analysis presented here leads to the conclusion that different types of algorithms may not make a large difference to the increase in computer speed required. We are left, therefore, with considerable freedom in the design of a special-purpose globular cluster machine, permitting the choice of an algorithm that will perform optimally for the most efficient type of hardware.

This work was supported in part by the NSF, under the US-Japan Cooperative Science Program.

1. Elson, R., Hut, P. & Inagaki, S. A. *Rev. Astr. Astrophys.* **25**, 565–601 (1987).
2. Spitzer, L. *Dynamical Evolution of Globular Clusters* (Princeton University Press, 1987).
3. *The Dynamics of Dense Stellar Systems* (ed. Merritt, D.) (Univ. of Cambridge Press, 1988).
4. Verbunt, F. in *Physics of Neutron Stars and Black Holes*, (Institute of Space and Astronautical Science, Tokyo, 1988).
5. Bettwieser, E. & Sugimoto, D. *Mon. Not. R. astr. Soc.* **208**, 439–509 (1984).
6. Raveché, H. J., Lawrie, D. H. & Despain, A. M. A *National Computing Initiative*, Proc. SIAM workshop, Virginia, 1987 (Soc. ind. appl. Math., Philadelphia, 1987).
7. Goodman, J. *Astrophys. J.* **313**, 576–595 (1987).
8. McMillan, S. L. W. & Lightman, A. P. *Astrophys. J.* **283**, 801–812 (1984); McMillan, S. L. W. & Lightman, A. P. *Astrophys. J.* **283**, 813–824 (1984).
9. McMillan, S. L. W. in *The Dynamics of Star Clusters* (eds Goodman, J. & Hut, P.) 331–333 (Reidel, Dordrecht, 1985).
10. McMillan, S. L. W. *Astrophys. J.* **307**, 126–138 (1986).
11. Cohn, H., Hut, P. & Wise, M. *Astrophys. J.* (submitted).
12. Aarseth, S. J. *Mon. Not. Roy. astr. Soc.* **126**, 223–255 (1963).

13. Aarseth, S. J. in *Multiple Times Scales* (eds Brackbill, J. U. & Cohen, B. I.) 377–418 (Academic, New York, 1985).
14. Makino, J. & Hut, P. *Astrophys. J. Suppl.* (in the press).
15. Ahmad, A. & Cohen, L. *J. comput. Phys.* **12**, 389–402 (1973).
16. Appel, A. *SIAM J. Sci. stat. Comput.* **6**, 85–103 (1985).
17. Porter, D. thesis, Univ. California (1985).
18. Jernigan, J. G. in *Dynamics of Star Clusters* IAU Symp. 113, (eds Goodman, J. & Hut, P.) 275–284 (Reidel, Dordrecht, 1985).
19. Barnes, J. & Hut, P. *Nature* **324**, 446–449 (1986).
20. Greengard, L. & Rokhlin, V. *J. comput. Phys.* **73**, 325–348 (1987).
21. Zhao, F. *Res. Rep. AI-TR-995*, Massachusetts Institute of Technology (1987).
22. Hénon, M. in *The Gravitational N-body Problem* IAU Coll. 10 (ed. Lecar, M.) 406–422 (Reidel, Dordrecht, 1972).
23. Heggie, D. C. & Mathieu, R. D. in *The Use of Supercomputers in Stellar Dynamics* (eds Hut, P. & McMillan, S. L. W.) 233 (Springer, New York, 1986).
24. Applegate, J. H. *et al. IEEE Transactions on Computers*, **c-34**, 822–831 (1985).

Spike–nucleocapsid interaction in Semliki Forest virus reconstructed using network antibodies

David J. T. Vaux, Ari Helenius & Ira Mellman*

Department of Cell Biology, Yale University School of Medicine, 333 Cedar Street, PO Box 3333, New Haven, Connecticut 06510, USA

Using in vitro immunization, we have reconstructed three consecutive steps of an idiotype network to show that the nucleocapsid of Semliki Forest virus contains a specific 'receptor' for the cytoplasmic tail of the E2 spike glycoprotein. This interaction could be the basis for the highly selective inclusion of viral glycoproteins—and exclusion of host cell surface proteins—during virus budding.

THE budding of enveloped animal viruses from infected cells is characterized by the selective inclusion of the viral genome and its accessory proteins^{1–3}. The budding mechanism is presently best understood for Semliki Forest virus (SFV), an alphavirus^{4–6}. The first step of SFV assembly is the packaging of a 42S single-stranded genomic RNA into an icosahedral nucleocapsid containing 180 copies of a single capsid protein. The capsids attach to the plasma membrane of the infected cell and progressively associate with 240 copies of the heterotrimeric (E1, E2, E3) spike glycoprotein complex during formation of the nascent virion. Host cell proteins are almost completely excluded during this process. Finally, the virus is released by pinching off from the plasma membrane.

Although the mechanism for the selective inclusion of viral protein into the budding membrane remains uncertain, a specific interaction between the assembled nucleocapsid and the cytoplasmic domain of the spike glycoprotein complex seems likely⁷. Chemical cross-linking⁸ and detergent extraction of intact virus particles⁹ have provided evidence for a physical association between the capsid and spike proteins. Electron microscopy has revealed 80 regularly spaced depressions in the surface of the nucleocapsids, possibly corresponding to the putative binding sites for spike glycoprotein trimers¹⁰.

To characterize the spike–capsid interaction further, we have used the structural mimicry of internal image anti-idiotype antibodies (see ref. 11 for review). If the E2 cytoplasmic tail contains a determinant for binding to a specific site on the nucleocapsid, it should be possible to generate antibodies which would recognize that determinant. Anti-idiotypes to the anti-E2 tail antibodies would be expected to recognize the E2 binding site on the capsid. Finally, if anti-anti-idiotypes were made, some of these should bind to the original antigen—the E2 cytoplasmic domain. Starting with a synthetic peptide corresponding to the E2 cytoplasmic tail, we found that all three steps of this antibody network—the idiotype, the anti-idiotype and the anti-anti-idiotype—could be produced as monoclonal antibodies by immunization *in vitro*. This showed that there is a specific interaction between the cytoplasmic domain of the E2 glycoprotein and the cytoplasmic nucleocapsid.

Polyclonal idiotype response production

As the SFV E2 glycoprotein is the only spike protein component with a significant cytoplasmic domain, we synthesized a peptide corresponding to its 31-residue cytoplasmic tail (NH₂-R-S-K-C-L-T-P-Y-A-L-T-P-G-A-A-V-P-W-T-L-G-I-L-C-C-A-P-R-A-H-A-COOH). To demonstrate its immunogenicity in the BALB/c mouse, conventional polyclonal antibodies were prepared first by repeatedly immunizing BALB/c mice with 25 µg of the unconjugated peptide. The resulting antisera reacted specifically

with the peptide as well as the intact E2 spike glycoprotein and its uncleaved precursor (p62) by immunoprecipitation (Fig. 1). The antisera were also weakly positive by indirect immunofluorescence (not shown).

To generate idiotype network antibodies, we used paired immunizations performed *in vitro*. Although anti-idiotype antibodies have been produced by conventional immunization procedures *in vivo*¹¹, *in vitro* immunization offers several advantages. The antigen is delivered directly to the immune cells, reducing the risk of antigen sequestration or proteolysis. The response to antigen is a primary one, resulting in immunoglobulin (Ig) M production, and occurs before antigen-specific T suppression^{12,13}. The antigen should therefore elicit responses from every relevant member of the B-cell repertoire, unaffected by the T-cell related tolerance effects often seen *in vivo*. Thus, an *in vitro* polyclonal response to an antigen should be an optimal source of the most complete 'library' of idiotype antibodies against *all* of the immunogenic epitopes of a given antigen. A further advantage is that syngeneic, second-round immunizations to generate anti-idiotype antibodies can be performed using the crude supernatant from the first *in vitro* immunization as antigen. This eliminates the need for identification and purification of a relevant monospecific idiotype before anti-idiotype production.

Unconjugated E2 peptide was used for immunization *in vitro* and, at the end of the five-day immunization period, immunofluorescence on SFV-injected cells was used to show that the culture supernatant was weakly positive. To confirm that anti-peptide antibodies were indeed produced, non-adherent cells were then used to generate hybridomas that were screened by immunofluorescence and an enzyme-linked immunosorbent assay (ELISA) against the starting peptide. Several positive clones were identified and a stable cell line secreting an anti-E2 IgM was isolated.

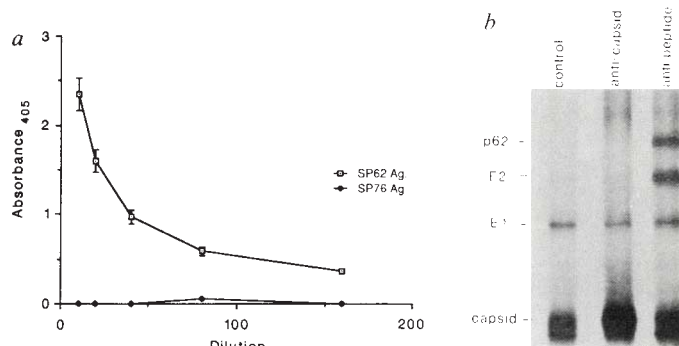
Monoclonal anti-idiotype preparation

To produce anti-E2 peptide anti-idiotypes, the complete culture supernatant from the first round immunization was concentrated, dialysed to remove any remaining peptide (determined using ¹²⁵I-labelled peptide) and added directly to a second culture of naive, syngeneic spleen cells. After five days, non-adherent cells were collected and fused with the SP-2 myeloma cell line. Wells containing hybridomas were first screened for reactivity by indirect immunofluorescence on uninfected and SFV-infected baby hamster kidney (BHK) cells (see below). Of the 120 wells initially plated after the fusion, 110 contained viable hybridomas and five exhibited virus-specific reactivity; each of these yielded stable, IgM-secreting hybridoma cell lines after repeated cloning in agarose. The results described below were obtained with the antibody secreted by a hybridoma line designated F13.

* To whom correspondence should be addressed.

Fig. 1 Reactivity of anti-cytoplasmic domain antibodies. *a*, ELISA of mouse anti-E2 cytoplasmic domain antiserum on the synthetic E2 tail peptide (open squares) and on a control synthetic peptide (closed squares). *b*, Immunoprecipitation of capsid, E2 and p62 from metabolically [35 S]methionine-labelled lysates of SFV-infected BHK-21 cells.

Methods. Synthetic peptides were purified by preparative HPLC and their identity confirmed by amino-acid analysis. For ELISA, solid-phase substrates were prepared by drying 0.5 μ g peptide in distilled water on 96-well plates (Nunc Immulon) overnight at 42 °C. Other substrates were plated at 50 μ l per well of 10–100 μ g ml $^{-1}$ protein in a 0.1 M sodium carbonate buffer at pH 8.5 and incubated overnight at 4 °C. Nonspecific sites were blocked by pre-incubation with PBS (Ca $^{2+}$ /Mg $^{2+}$ -free) containing 10% normal goat serum for 2 h at room temperature. Wells were washed with PBS and incubated with the first antibody diluted in the blocking solution (50 μ l per well) for 2 h at room temperature. The plates were washed in five changes of PBS over 30 min and then re-blocked for 30 min. Alkaline phosphatase-conjugated, affinity-purified goat second antibodies (Zymed) were then added at 50 μ l per well at dilutions from 1:300 to 1:800 in PBS containing 10% goat serum. After 2 h, the plates were washed and then incubated with phosphatase substrate (disodium p-nitrophenyl phosphate; Sigma) in 50 mM sodium glycinate pH 8.7 containing 4 mM MgCl $_2$ at room temperature. Plates were read in an automatic plate scanner after 2 h. The result of a representative experiment is shown as means of triplicate determinations with error bars to plus and minus 1 standard deviation. Control peptides included a 22-residue synthetic peptide corresponding to part of the LDL receptor cytoplasmic domain (NH $_2$ -K-N-W-R-L-K-N-I-N-S-I-N-F-D-N-P-V-Y-Q-K-T-T-COOH; closed squares) and two water-soluble synthetic peptides corresponding to parts of the ectodomain of the human Fc receptor (ref. 30; data not shown). Specific polyclonal antisera raised against the control peptides were used as positive controls in these control experiments to check adequate peptide retention by the ELISA plates. For the immunoprecipitation experiments, BHK-21 cells were plated to 70% confluence in 100-mm tissue culture dishes and infected the following day with SFV at a multiplicity of infection (MOI) of 500. After 2 h at 37 °C, the plates were washed twice in RPMI and incubated for 30 min at 37 °C in methionine-free medium containing 2 μ g ml $^{-1}$ actinomycin D. This medium was then removed and replaced with 5 ml of fresh medium containing 0.5 mCi [35 S]methionine. The plates were incubated at 37 °C for 150 min, washed in PBS, and collected by scraping in 4 ml per dish of cold homogenization buffer (PBS with 300 mM NaCl, 2 mM EDTA, 20 mM NaN $_3$, 1 mM phenylmethylsulphonylfluoride, 0.22 U ml $^{-1}$ aprotinin and 200 μ g ml $^{-1}$ bovine serum albumin (BSA), pH 8.2). The cells were homogenized by eight strokes of a stainless steel Dounce homogenizer on ice and centrifuged in a Beckman microfuge at 4 °C for 5 min. The postnuclear homogenate was stored at –70 °C until used, then thawed and brought to 1% with NP-40. The lysate was cleared by a 2 h incubation with 5% (w/v) washed fixed *Staphylococcus aureus* on a microshaker at 4 °C and microfuged for 10 min. Aliquots (50 μ l) were incubated with 3–10 μ l of mouse or rabbit polyclonal antisera at 4 °C in an Eppendorf microshaker. After 2 h, 20 μ g of a goat anti-mouse IgG was added to precipitations involving murine first antibodies. After another 2 h, 50 μ l of a 10% (w/v) suspension of *Staph. aureus* was added to all samples. Incubation was continued for 2 h and the immunoadsorbant was washed three times in 0.6 M NaCl, 10 mM Tris-HCl pH 7.2 at 4 °C. Finally, the pellets were washed once in distilled water and resuspended in sample buffer containing 5 μ g cold SFV per sample, heated to 95 °C for 5 min and run unreduced on 10% polyacrylamide gels. Gels were stained with Coomassie blue, neutralized in PBS, and incubated in salicylate (Chamberlain, 1979) before drying and autoradiography (from 12 h to 5 days). Under these conditions, minor nonspecific precipitations of E1 and capsid protein are seen in all lanes, whereas only the anti-cytoplasmic domain antibody precipitates any E2 or p62.



As shown in Fig. 2, F13 IgM exhibited a punctate immunofluorescence pattern throughout the cytoplasm of SFV-infected BHK-21 cells (Fig. 2*b*). The pattern was distinct from that obtained with a rabbit anti-SFV spike glycoprotein antiserum which had a characteristic staining pattern of the endoplasmic reticulum and Golgi region on methanol-fixed and permeabilized cells (Fig. 2*a*). It was, however, similar to the distribution of nucleocapsids visualized by a rabbit polyclonal anti-capsid antiserum (Fig. 2*c*). The F13 staining was known to be due to virus infection, as no staining was observed on mock-infected cells (Fig. 2*d*) or an SFV-infected cells not permeabilized before antibody addition, indicating that the antigen was intracellular.

The F13 antigen

For biochemical identification of the antigen, the F13 antibody was used for immunoprecipitation from [35 S]methionine-labelled, SFV-infected cells. In preliminary experiments, the use of ELISA showed that the antigen was sensitive to low concentrations of non-ionic detergent (for example 0.1% Triton X-100, Triton X-114, β -D-octylglucoside, CHAPS, Tween 20). Therefore, immunoprecipitation was performed by first allowing the antibody to bind to the antigen in crude post-nuclear supernatants before detergent solubilization for the immunoadsorption and washing steps. As shown in Fig. 3, F13 IgM precipitated the capsid protein; no other viral or host cell proteins were immunoprecipitated. Although there was some nonspecific sticking, insignificant amounts of capsid were precipitated by control monoclonal IgMs. This background precipitation was not concentration-dependent and occurred if the control IgM was omitted entirely (Fig. 3). F13 was also found by ELISA not

to react with the E2 cytoplasmic domain peptide (see below).

These results indicated that F13 recognized a detergent-sensitive epitope on the SFV nucleocapsid and did not recognize a host cell antigen or the starting synthetic peptide. Thus, it was likely that F13 arose as an internal image anti-idiotypic against the combining site of an antibody against the cytoplasmic domain peptide.

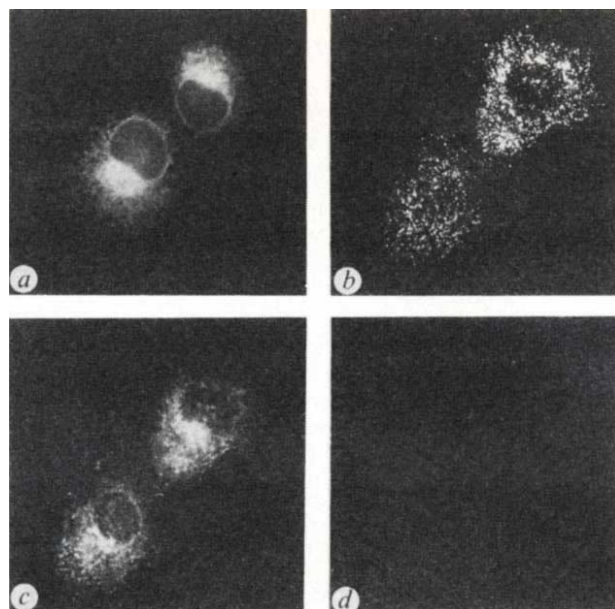
Anti-anti-idiotypic production

The formal possibility remained that the F13 hybridoma represented the immortalization of an irrelevant B-cell specificity from the naive BALB/c spleen which happened to react with an epitope on the SFV capsid. To exclude this, we used F13 cells as antigen for preparation of anti-anti-idiotypic monoclonals. *In vivo* immunization was used to obtain IgG antibodies to screen for binding to F13 IgM using γ -chain-specific second reagents.

Syngeneic BALB/c mice were immunized and boosted intravenously once with washed, glutaraldehyde-fixed F13 cells. They received an intravenous boost of purified F13 IgM four days before fusion. The resulting hybridoma supernatants were tested for their ability to recognize the starting F13 IgM and the E2 cytoplasmic domain peptide. Results obtained for the 16 highest-binding supernatants are shown in Fig. 4*a*. There were two classes of positive reactivity. Well 14 is an example of the first class: strong binding to F13 IgM (Fig. 4*a*, top) but no reactivity to the starting cytoplasmic domain peptide (Fig. 4*a*, bottom). This pattern of reactivity was consistent with 'framework' anti-idiotypes¹¹: antibodies which detect novel epitopes on the immunizing antibody outside the antigen combining site.

Fig. 2 Intracellular localization of viral proteins in infected cells. SFV-infected (*a*, *b* and *c*) or mock-infected (*d*) BHK-21 cells were fixed, permeabilized and labelled with a rabbit polyclonal antibody to SFV spike glycoproteins (*a*), the monoclonal anti-idiotypic antibody F13 IgM (panels *b* and *d*), or a rabbit polyclonal antibody against isolated SFV nucleocapsids (panel *c*). No F13 reactivity was seen if infected cells were stained without permeabilization (unfixed cells at 4 °C or cells fixed with 3% (w/v) paraformaldehyde without subsequent methanol treatment). The apparent lack of cell surface staining in *a* reflects the fixation used, because methanol decreases the amount of spike protein surface fluorescence. *b* and *c* were obtained by double immunofluorescence and represent the same field exposed for fluorescein and rhodamine, respectively.

Methods. To prepare thymus-conditioned medium (TCM), single cell suspensions from the thymuses of three-week-old BALB/c and C57/B1 mice were co-cultured at a density of 2×10^6 thymocytes of each strain per ml in a medium consisting of alpha-modified Eagle's medium (MEM) with 50 μ M 2-mercaptoethanol and 2% type-100 rabbit serum (Diagnostic Biochemistry). After 48 h, the supernatant was clarified by centrifugation and 0.22 μ m filtered before storage at -20 °C. The TCM retained the ability to support *in vitro* immunization for at least six months when prepared in this way. *In vitro* immunizations were performed by a modification³⁰ of the method of Borrebaeck and Moller³¹. Briefly, the spleens were dissected from two 12-week-old BALB/c mice and a single-cell suspension was prepared by teasing the organs apart with two 18-gauge needles. The cells were washed twice in alpha-MEM containing 50 μ M 2-mercaptoethanol and resuspended to a final volume of 30 ml in 50% TCM and 50% normal lymphocyte medium (NLM) (alpha-MEM, 2 mM sodium pyruvate, 50 μ M 2-mercaptoethanol, 1 $\times$ Gibco non-essential amino acids, 10 mM HEPES pH 7.2 and 20% heat inactivated fetal calf serum (JR Scientific)). The cell density was approximately 1×10^7 nucleated cells per ml. 10 μ g of sterile-filtered E2 cytoplasmic domain peptide in 0.1 ml was then added and the mixture cultured in a T-75 flask at 37 °C in 5% CO₂ without disturbance for five days. Non-adherent cells were then released by brick shaking and collected by centrifugation at 400g for 5 min. The cells were used for hybridoma production by fusion with SP2/0-Ag14 myeloma cells (refs 32, 33 and Fig. 4 legend). The supernatant from this *in vitro* immunization was concentrated 15-fold by dialysis against solid polyethylene glycol (M_r , 20,000), dialysed extensively against PBS and used as the antigen for an identical repeat *in vitro* immunization. The non-adherent cells from this second immunization were fused with SP2 myeloma cells and the resulting hybridomas screened by indirect immunofluorescence. BHK-21 cells were plated at 80% confluence on multi-well slides and infected at an MOI of 10–100 with SFV in RPMI supplemented with 1 mg ml⁻¹ BSA and 10 mM HEPES (pH 6.8) for 2 h, washed and incubated in growth medium for 3 h. Mock-infected cells were incubated in the same buffer without virus. The cells were washed with PBS and fixed with methanol for 6 min at -20 °C. Nonspecific sites were blocked with PBS containing 10% goat serum for 30 min and the wells were incubated with first antibody for 1 h. For double-label immunofluorescence, both first antibodies were applied together (20 μ g ml⁻¹ for purified F13 IgM). The slides were washed and reblocked for 30 min. Fluorochrome-conjugated second antibodies (Zymed) diluted 1:40 in PBS with 10% goat serum were then added for 1 h. After washing in PBS and water, the slides were mounted in Moviol and photographed with Tri-X Pan on a Zeiss Photomicroscope III equipped with epifluorescence illumination.



In contrast, wells 3, 6 and 9 contained antibodies with strong reactivity to both purified F13 IgM (Fig. 4*a*, top) and cytoplasmic domain peptide (Fig. 4*a*, bottom). None of these antibodies gave signals above background when assayed by ELISA on bovine serum albumin (BSA) or a range of other water-soluble peptides of approximately similar size and composition, including a peptide corresponding to the cytoplasmic domain of the low-density lipoprotein receptor (data not shown). This pattern of reactivity was consistent with 'internal image' anti-idiotypes¹¹. A representative IgG-secreting hybridoma was cloned (well 9) and designated 3G10. Purified 3G10 IgG bound to the cytoplasmic domain peptide (Fig. 4*b*, bottom). As expected, the anti-idiotypic F13 IgM with anti-capsid reactivity did not bind to the E2 tail peptide (Fig. 4*b*, middle).

The anti-idiotypic 3G10 IgG not only recognized the synthetic peptide by ELISA, but also labelled permeabilized SFV-infected BHK-21 cells. Although its distribution was less extensive than that of the polyclonal antisera to the spike glycoprotein complex (Fig. 2*c*), it was similar to the weaker pattern observed using the murine anti-peptide antisera (not shown) or a rabbit polyclonal antiserum against the same synthetic peptide¹⁴.

An immunofluorescence assay has been used to confirm that, as an internal image anti-idiotypic to F13 IgM, 3G10 IgG also recognizes the antigen combining site on F13 IgM and competes for binding to the SFV nucleocapsid. When F13 IgM was pre-incubated with excess goat serum (Fig. 5*a*) or monoclonal mouse IgG against an irrelevant epitope (Fig. 5*b*), the normal F13

pattern and intensity was observed. However, if F13 was pre-incubated with purified 3G10 IgG, there was a marked loss of the usual F13 staining pattern when this mixture was used to stain SFV-infected cells (Fig. 5*c*).

Structures bearing the F13 epitope

The F13 determinant reflects a functional domain on the nucleocapsid involved in its interaction with viral spike glycoproteins and thus the intracellular distribution of capsids being the F13 determinant may be related to the mechanism of virus assembly. Nucleocapsids are overproduced during alphavirus infections with only a fraction of the most recently synthesized capsids being incorporated into new virions¹⁵; thus, it seems likely that the F13-positive nucleocapsids include the functionally active subset. The relationship between the F13-positive capsids and the total capsid population was examined by double-label immunofluorescence. Five hours after infection, BHK-21 cells were stained with F13 IgM (detected using fluorescein isothiocyanate-(FITC)-conjugated second antibody) and a rabbit anti-capsid antiserum (visualized using tetramethylrhodamine isothiocyanate (TRITC)-conjugated second antibody). Superimposition of the two images clearly showed that almost all of the F13-positive structures (green) were also stained by the anti-capsid antibodies (red) and thus appeared as yellow spots (Fig. 6). As expected, a majority of the structures were reactive only with the anti-capsid antibody, reflecting the accumulating pool of presumably aberrant nucleocapsids which do not form virions.

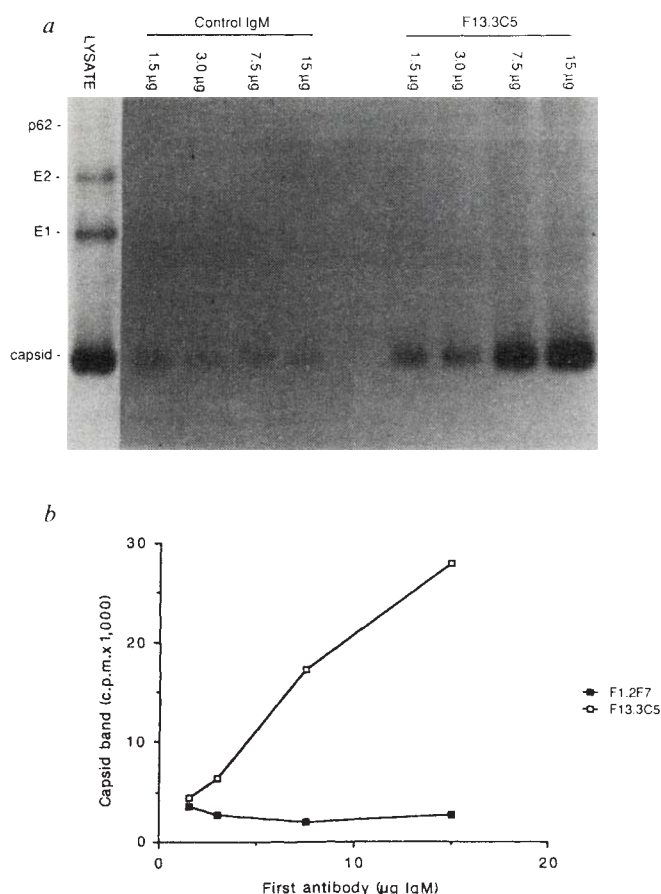


Fig. 3 F13 IgM specifically immunoprecipitates the SFV capsid protein. *a*, Autoradiograph of immunoprecipitation from metabolically ^{35}S -labelled, SFV-infected, BHK-21 cell homogenates using increasing amounts of first antibody. *b*, Quantitation of F13 capsid immunoprecipitation.

Methods. Radiolabelled homogenates (prepared as in Fig. 1 legend) were incubated in aliquots of 50 μl . With a range of identical doses of an irrelevant mouse IgM monoclonal antibody (*a*, left) or F13 IgM (*a*, right; clone F13.3C5). All incubations were at 4°C on an Eppendorf microshaker. After 2 h to allow antibody-antigen interaction to occur, NP-40 was added to a final concentration of 1% and the lysates were cleared with fixed *Staph. aureus*. The cleared lysates containing pre-bound first antibody were incubated for 2 h with 20 μg per sample of goat anti-mouse IgG or IgM as appropriate and then treated exactly the same as immunoprecipitations from lysates (see Fig. 1 legend). F13 recognizes the capsid protein in a dose-dependent way and does not recognize any other cellular or viral protein (labelled SFV structural proteins are shown in the lane at left). Labelled capsid bands from the gel autoradiographed in *a* were excised and counted by liquid scintillation spectroscopy. Background radioactivity, determined using gel strips of the same size from empty lanes, contained less than 100 c.p.m.

Discussion

We have used internal image anti-idiotypic antibodies to show that there is a specific receptor-ligand-like interaction between the nucleocapsid and the cytoplasmic domain of the E2 spike glycoprotein of Semliki Forest virus. This interaction, originally predicted by Simons and Garoff^{7,8}, is likely to be critical in the organization of the budding of SFV and related viruses from infected cells. Our results support a mechanism whereby newly formed nucleocapsids attach to the cytoplasmic face of the plasma membrane by recognizing the E2 endodomain and recruit additional spike glycoprotein complexes into the nascent bud. Host cell membrane proteins, whose cytoplasmic domains lack the determinant involved in nucleocapsid binding, would not be recruited; in fact, they might be sterically excluded as a

result of the close association of the spike glycoprotein trimers^{2,10}. The spike-nucleocapsid interaction is likely to reflect the mechanism of virus budding, but its possible role in other functions—such as the stability or infectivity of the virus particle—cannot be excluded.

These results also demonstrate that it is possible to use internal image anti-idiotypic antibodies to reconstruct interactions between membrane protein 'ligands' and potential cytoplasmic 'receptors' which may specify membrane traffic events. Moreover, the development of a method to produce monoclonal anti-idiotypes by *in vitro* immunization provides a rapid and practical new approach to a wide variety of problems related to protein interactions and intracellular recognition.

Production of monoclonal anti-idiotypes *in vitro*. Although the SFV E2 cytoplasmic domain only contains 31 amino acids, it is long enough to have a number of distinct antigenic epitopes, only one of which can be the site for capsid binding. No assay exists for the identification of the relevant epitope. The generation of an idiotype library of anti-peptide antibodies which was assumed to contain antibodies against the relevant epitope overcame this problem. The library was then used as the antigen for a second *in vitro* immunization, resulting in the production of monoclonal anti-idiotypic antibodies (for example F13 IgM), the functionally relevant determinant. Critical to the process, however, was the confirmation that F13 IgM represented an authentic internal image anti-idiotypic by isolating monoclonal anti-F13 antibodies and showing that a subset of these were specifically reactive with both F13 IgM and the starting cytoplasmic domain peptide. These monoclonal antibodies were thus 'anti-anti-idiotypic' with respect to the starting peptide and could only have arisen if F13 was a true internal image anti-idiotypic. Additionally, these internal image anti-anti-idiotypes (for example 3G10 IgG) must recognize the specific epitope within the peptide that interacts with the nucleocapsid. The final result is monoclonal antibodies against the active sites of both the nucleocapsid 'receptor' (F13 IgM) and the E2 tail 'ligand' (3G10 IgG).

Although the SFV spike is a heterotrimer containing three copies of the E2 glycoprotein, F13 IgM was produced starting with a synthetic E2 tail peptide which is not likely to have any oligomeric organization. Thus, the relevant signal specifying E2 tail-capsid interaction must be present on the synthetic peptide in aqueous solution. This implies that the trimeric state of the E2 tail in the spike glycoprotein complex cannot be essential for expression of this signal, although it is possible that positive cooperativity between three sites enhances its effectiveness *in vivo*.

Conservation of the spike-nucleocapsid signal. As the E2-nucleocapsid interaction probably reflects a functionally important recognition step in virus budding, we expected that the F13 determinant might be conserved on the capsids of alphaviruses other than SFV. Indeed, strong labelling was seen in BHK-21 cells infected with each of 14 different alphaviruses tested (including Sindbis, Ross river, Eastern equine encephalitis, Whartaroa and O'Nyong Nyong viruses; D.J.T.V. *et al.*, manuscript in preparation). This reactivity was also observed in cells infected with viruses from the related flavivirus group (Japanese encephalitis, West Nile, Tembusu and Banzi viruses). Thus, the nucleocapsid determinant recognized by F13 IgM seems to be conserved through virus evolution. However, the structural organization and budding of flaviviruses is more complex (and less well understood) than alphaviruses. At least some flaviviruses have a matrix protein intercalated between the nucleocapsid and spike-containing lipid envelope¹⁶. It is not known whether the site recognized by F13 IgM in flaviviruses interacts with the spike glycoprotein itself or with an intervening matrix protein.

Although the extent of the conservation of the F13 epitope is not yet known, it is not a pan-viral structure because it is not expressed in cells infected with influenza virus (an

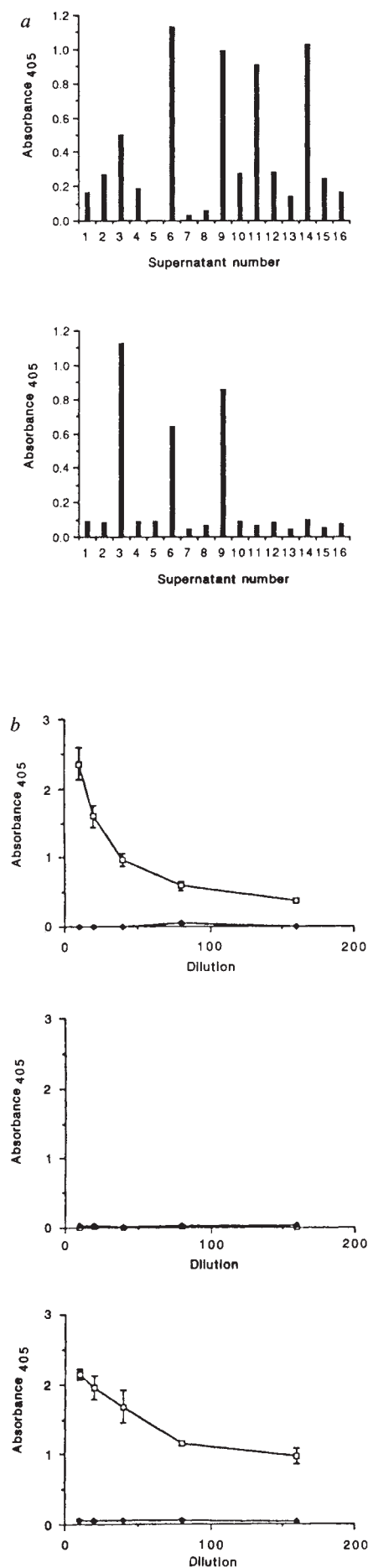


Fig. 4 Identification of anti-F13 anti-idiotype monoclonal antibodies. *a*, Screening for anti-F13 anti-idiotype reactivity. Supernatants from hybridoma-containing wells were assayed by ELISA using purified F13 IgM (top) and synthetic E2 cytoplasmic domain peptide (bottom) as solid-phase antigen. Wells 3, 6 and 9 contained antibodies which reacted with both the antigen (F13 IgM) and the starting peptide whereas others (wells 11 and 14) contained antibodies that recognized only F13 IgM. The positive control (mouse anti-peptide antiserum at 1:200 dilution) had a mean absorbance of 2.5 and the blank (medium only) had a mean absorbance of 0.07. *b*, Reactivity of the anti-peptide idiotype, anti-idiotype and anti-anti-idiotype monoclonal antibodies. The reactivity of the anti-E2 cytoplasmic domain peptide mouse polyclonal antiserum (top), the monoclonal anti-peptide anti-idiotype F13 IgM (middle) and the monoclonal anti-anti-idiotype 3G10 IgG (bottom) against the starting E2 peptide was determined by solid-phase ELISA. Open squares indicate reactivity to the synthetic E2 cytoplasmic domain peptide and closed squares indicate reactivity to an irrelevant peptide (see Fig. 1 legend).

Methods. Hybridomas were produced from the spleens of mice hyper-immunized with purified F13 IgM and fixed F13 cells as described³³. Feeder layers consisting of syngeneic mouse total peritoneal exudate cells at 1×10^5 per well in 0.3 ml NLM were established in 24-well plates. The fusion mixture was plated at 1×10^5 spleen input cell per well in 0.3 ml NLM containing 100 μ M (1 $\times$) hypoxanthine and 2 μ g ml⁻¹ (1 $\times$) azaserine. After 24 h, the wells were fed with 0.3 ml $\times$ NLM containing 2 $\times$ hypoxanthine and 2 $\times$ azaserine and on subsequent days with NLM containing 1 $\times$ hypoxanthine.

orthomyxovirus) or vesicular stomatitis virus (a rhabdovirus). These results are consistent with previous observations that pseudotypes may form between two alphaviruses¹⁷ or between an alphavirus and a flavivirus¹⁸ but not between an alphavirus nucleocapsid and vesicular stomatitis virus glycoprotein G¹⁹. The mechanism of alphavirus budding, however, is not directly applicable to all enveloped viruses. Retroviruses such as Rous sarcoma virus, whose budding mechanism is poorly understood and probably very different from the alphaviruses, are able to produce virions using mutant forms of the *env* spike glycoprotein that lack a cytoplasmic tail²⁰. Presumably other proteins (such as *gag*) participate in interactions which organize newly formed viral components during budding.

Conventional antibodies to nucleocapsids of alphaviruses frequently show little, if any, cross-reactivity. For example, the rabbit antiserum against SFV capsid protein did not recognize the capsids of any of the other alphaviruses tested. In contrast, the broad cross-reactivity of F13 IgM probably results from its specificity for an epitope in the spike binding site. Functionally important sites on proteins are often highly conserved and are therefore good targets for cross-reactive antibodies. They are consequently good candidates as vaccine or diagnostic antigens. The anti-idiotype route allows selection of antibodies to functional sites on proteins against which it may be difficult to obtain a strong immune response by conventional immunization. The ability of internal image anti-idiotype antibodies to identify such sites within viruses makes them potentially powerful tools in vaccine development.

Use of internal image anti-idiotype antibodies. Mimicry of receptor structure by anti-ligand anti-idiotypes was first demonstrated by Sege and Peterson using retinol binding protein as the ligand and human pre-albumin as the receptor²¹. Other groups have used anti-ligand anti-idiotype antibodies to probe receptor-ligand interactions at the cell surface. These include identification of the mammalian reovirus type 3 receptor^{22,23}, generation of antibodies to the acetylcholine receptor²⁴, insulin receptor²⁵, thyroid stimulating hormone (TSH) receptor²⁶, lymphocyte retroviral binding site²⁷ and the vasopressin receptor in rat brain²⁸. An anti-idiotype approach using polyclonal rabbit antisera has also been used successfully to identify a receptor apparently involved in protein import into chloroplasts²⁹.



Fig. 5 Anti-anti-idiotypic 3G10 competes with antigen for F13 binding. F13 IgM at $1 \mu\text{g ml}^{-1}$ was pre-incubated with $100 \mu\text{g ml}^{-1}$ of an irrelevant monoclonal mouse IgG (b) or $100 \mu\text{g ml}^{-1}$ 3G10 IgG (c) for 2 h and then used to stain methanol-fixed and permeabilized SFV infected BHK-21 cells. The pattern and intensity of labelling in b are indistinguishable from that in control wells stained with $1 \mu\text{g ml}^{-1}$ F13 without pre-incubation (a). Pre-incubation of F13 with 3G10 results in a marked loss of labelling (c). All three panels received identical exposures both during initial photography and subsequent printing.

Previous studies involving anti-idiotypic reagents to examine receptor-ligand-like interactions have required the identification and isolation of a single relevant idiotypic before anti-idiotypic production which is time consuming. Our approach does not require the identification or purification of ligand or receptor, nor does it require that a functionally relevant single idiotypic be identified; it also can produce useful anti-idiotypic reagents less than three weeks. In addition to the anti-viral antibodies described, we have used the approach to produce idiotypic and potential anti-idiotypic antibodies beginning with complex cellular organelles (for example lysosomes) as antigen, as well as with endogenous cellular proteins and derived synthetic peptides. By definition, internal image anti-idiotypes should recognize sites involved in intermolecular recognition events and so these antibodies should interfere with a functional interaction between the antigen ('receptor') and its presumptive 'ligand', be it another protein, organelle or viral component. Experiments are in progress to determine if the anti-idiotypic F13 and the anti-anti-idiotypic 3G10 interfere sufficiently with the SFV spike-capsid interaction to disrupt viral assembly and budding.

We thank Klaus-Peter Zimmer, Hans Stukenbrok, Jürgen

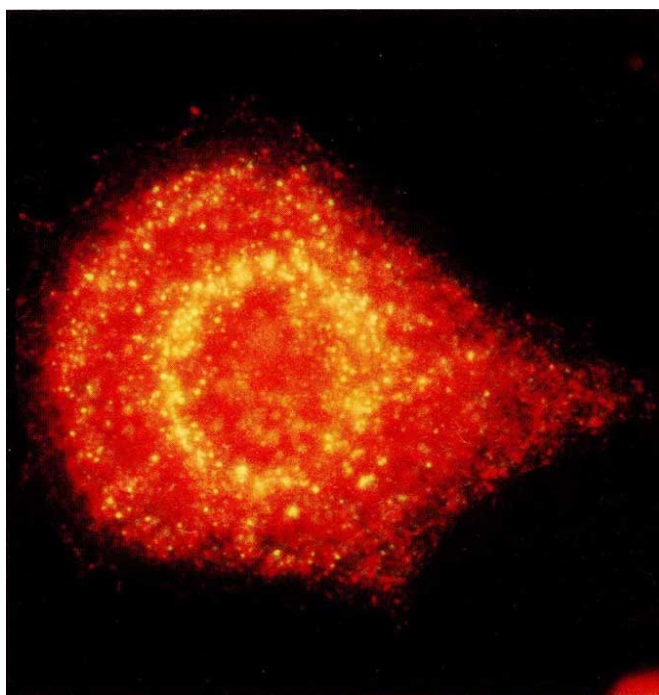


Fig. 6 Anti-idiotypic F13 detects a subset of nucleocapsids in SFV-infected BHK-21 cells. BHK-21 cells were infected for 5 h with SFV and then fixed and permeabilized with methanol. Double-label indirect immunofluorescence was then carried out using a polyclonal rabbit anti-nucleocapsid antiserum (rhodamine label) and the anti-idiotypic F13 IgM (fluorescein label). Cells were photographed by double exposure of single frames of colour slide film to assess the degree of co-localization of rhodamine- and fluorescein-positive structures. Structures staining yellow represent co-localization of both fluorochromes; some images were photographed as double exposures fractionally out of register to confirm the presence of both labels (not shown). Background levels for all antibody combinations were as shown for mock-infected cells (Fig. 2d), and there was no detectable cross-reactivity between second antibodies. Almost all fluorescein-staining F13 IgM-positive structures were contained within the more extensive population of structures labelling with the anti-capsid antibody. The few green fluorescein-staining structures probably resulted from out-of-register localization due to differences in optical properties associated with detecting fluorescein and rhodamine fluorescence.

Kartenbeck and Susan Froshauer for advice and assistance with electron microscopy and cell disruption; Andy Shyjan and Helen Plutner for advice and assistance with cell culture and ELISA assays; and Susan Froshauer and Margaret Kielian for helpful discussions. We are indebted to Robert Shope (for providing access to the virus stocks of the Yale Arbovirus Research Unit) and to Mark Marsh for rabbit polyclonal antisera against isolated SFV capsid protein and against the 29S complex containing viral spike glycoproteins. This work was supported by grants from the NIH (to IM and AH) and by awards from the Hull and Warner Cancer Research Foundations (to D.J.T.V.).

Received 11 May; accepted 30 August 1988.

1. Roman, L. M. & Garoff, H. *Trends biochem. Sci.* **119**, 428-432 (1985).
2. Simons, K. & Fuller, S. in *Biological Organisation: Macromolecular Interactions at High Resolution* (in the press).
3. Simons, K. & Warren, G. *Adv. Protein Chem.* **36**, 79-132 (1984).
4. Kääriäinen, L. & Söderlund, H. *Curr. Topics Microbiol. Immun.* **82**, 16-57 (1978).
5. Dubois-Dalq, M., Holmes, K. & Rentier, B. *Assembly of Enveloped Animal Viruses* (Springer, Vienna, 1984).
6. Schlesinger, M. & Schlesinger, S. *The Togaviridae and Flaviviridae* (Plenum, New York, 1986).
7. Simons, K. & Garoff, H. *J. gen. Virol.* **50**, 1-21 (1980).

8. Garoff, H. & Simons, K. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3988-3992 (1974).
9. Helenius, A. & Kartenbeck, J. *Eur. J. Biochem.* **106**, 613-618 (1980).
10. Fuller, S. D. *Cell* **48**, 923-934 (1987).
11. Gaulton, G. N. & Greene, M. L. *A. Rev. Immun.* **4**, 253-280 (1986).
12. Schreier, M. H. & Lefkowitz, I. *Immunology* **36**, 743-752 (1979).
13. Reading, C. L. *J. Immun. Meth.* **53**, 261-291 (1982).
14. Cutler, D. F., Melancon, P. & Garoff, H. *J. cell. Biol.* **102**, 902-910 (1986).
15. Scheele, C. H. & Pfefferkorn, E. R. *J. Virol.* **3**, 369-375 (1969).
16. Rice, C. M. *et al. Science* **229**, 726-733 (1985).
17. Burge, B. W. & Pfefferkorn, E. R. *Nature* **210**, 1937-1940 (1966).
18. Lagwinski, E., Stewart, C. C., Adles, C. & Schlesinger, S. *Virology* **65**, 204-214 (1975).
19. Zavadova, Z., Zavadova J. & Weiss, R. *J. gen. Virol.* **37**, 557-567 (1977).
20. Perez, L. G., Davis, G. L. & Hunter, E. *J. Virol.* **61**, 2981-2988 (1987).
21. Sege, K. & Peterson, P. A. *Nature* **271**, 167-169 (1978).
22. Sung Co, M., Gaulton, G. N., Fields, B. N. & Greene, M. I. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1494-1498 (1985).
23. Sung Co, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 5315-5318 (1985).
24. Erlanger, B. F. *et al. Immun. Rev.* **94**, 23-37 (1986).
25. Elias D., Cohen, I. R., Schechter, Y., Spier, Z. & Golander, A. *Diabetes* **36**, 348-354 (1987).
26. Farid, N. R. & Lo, T. C. Y. *Endocr. Rev.* **6**, 1-23 (1985).
27. Ardman, B., Khirya, R. H. & Schwartz, R. S. *J. exp. Med.* **161**, 669-678 (1985).
28. Knigge, K. M., Piekut, D. T. & Berlove, D. J. *Cell Tissue Res.* **246**, 509-513 (1986).
29. Pain, D., Kanwar, Y. S. & Blobel, G. *Nature* **331**, 232-237 (1988).
30. Stuart, S. G. *et al. J. exp. Med.* **166**, 1668-1684 (1987).
31. Borrebaeck, C. A. K. & Moller, S. A. *J. Immun.* **136**, 3710-3715 (1986).
32. Shulman, M., Wilde, C. & Kohler, G. *Nature* **276**, 269-270 (1978).
33. Mellman, I., Steinman, R. M., Unkeless, J. C. & Cohn, Z. A. *J. cell. Biol.* **86**, 712-722 (1980).

A peptide model of a protein folding intermediate

Terrence G. Oas & Peter S. Kim

Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA
Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

It is difficult to determine the structures of protein folding intermediates because folding is a highly cooperative process. A disulphide-bonded peptide pair, designed to mimic the first crucial intermediate in the folding of bovine pancreatic trypsin inhibitor, contains secondary and tertiary structure similar to that found in the native protein. Peptide models like this circumvent the problem of cooperativity and permit characterization of structures of folding intermediates.

ONE of the basic goals of protein folding studies is to determine the structures of folding intermediates, ultimately at atomic resolution, and to understand the thermodynamic basis of their formation. It is difficult to obtain folding intermediates of single-domain proteins at equilibrium because the cooperativity of folding is high. Folding intermediates can be populated kinetically but they are short-lived, so detailed structural analyses (using nuclear magnetic resonance spectroscopy or X-ray crystallography for example) are difficult. Amide proton labelling methods have provided insights into the structures of kinetic folding intermediates^{1,2}, but a high-resolution structure has not yet been determined for any folding intermediate of a single-domain protein.

The most well characterized protein folding pathway is described in terms of disulphide-bond formation in the oxidative refolding of bovine pancreatic trypsin inhibitor^{3,4} (BPTI). Reduction of the three disulphide bonds in native BPTI unfolds the protein in the absence of denaturants. The pathway of disulphide-bond formation in the refolding of reduced BPTI has been determined by Creighton (for review, see ref. 5). The term 'pathway' has led to some confusion and we use it here in the same sense as Creighton: a pathway describes the kinetically most accessible routes for folding. Other routes for folding are possible but are not depicted because they are less probable.

A simplified version of the pathway for folding of BPTI (ref. 3) is shown in Fig. 1a. A crucial intermediate in this pathway contains a single disulphide bond between residues 30 and 51 (ref. 6) and is called [30-51]. Because disulphide-bond formation and protein folding are thermodynamically linked processes, the conformations that favour particular disulphide bonds in folding are stabilized by those disulphides³. Thus, structural characterization of intermediates trapped by blocking the remaining thiols may allow one to understand the pathway in structural terms, provided that the blocking groups do not alter the conformations of these intermediates. Unfortunately, solubility limitations and the presence of substantial unfolded regions have hindered detailed structural characterization of trapped BPTI folding intermediates. Nevertheless, several trapped intermediates have been surveyed using NMR⁷: spectra of the intermediates were found to resemble spectra of BPTI (ref.

8) more closely as the number of native disulphide bonds in the intermediates increased. A comparison of the resonance pattern in the NMR spectrum of [30-51] with that of native BPTI suggested⁷ that at low temperatures [30-51] contains part of the central β -sheet found in BPTI.

We have designed and synthesized a small (30 residues) synthetic analogue of [30-51]. The analogue (called $\alpha\alpha\beta$) is a disulphide-bonded peptide pair in which two short peptides are connected by a disulphide bond corresponding to the 30-51 disulphide of BPTI (Fig. 1b). $\alpha\alpha\beta$ is very soluble in aqueous solution and does not aggregate. As judged by circular dichroism (CD) and NMR, $\alpha\alpha\beta$ is >90% folded in aqueous solution (pH 6) at 4 °C. The structure unfolds when the disulphide bond is reduced. Two-dimensional nuclear Overhauser effect spectroscopy (NOESY) indicates that $\alpha\alpha\beta$ contains much of the secondary and tertiary structure present in the corresponding region of native BPTI. These results indicate that native-like structure can form early in protein folding, and that peptide models can be used to characterize the structures of transient folding intermediates. In future work, it may be possible to crystallize peptide models like $\alpha\alpha\beta$.

Design of a synthetic analogue of [30-51]

Creighton's pathway³ for the folding of BPTI, also referred to as the rearrangement pathway, is shown in Fig. 1a. BPTI can refold into a native-like structure without the 30-51 disulphide⁹, but the resultant species is a non-productive intermediate that must undergo unfavourable disulphide rearrangements to complete folding⁴. This non-productive intermediate [N(30SH, 51SH)] and minor species have been omitted from Fig. 1a (see Fig. 6 of ref. 4 for a more complete description). Although there are fifteen possible one-disulphide combinations, [30-51] accounts for 60% of the one-disulphide species observed. The importance of [30-51] is clear from the observation that all further intermediates in the pathway retain the 30-51 disulphide (Fig. 1a).

Of the four remaining cysteines in [30-51], residues 5, 14 and 38 participate in formation of the second disulphide bond with approximately equal probability. In contrast, Cys 55 does not participate to an appreciable extent^{3,10}. None of the pre-

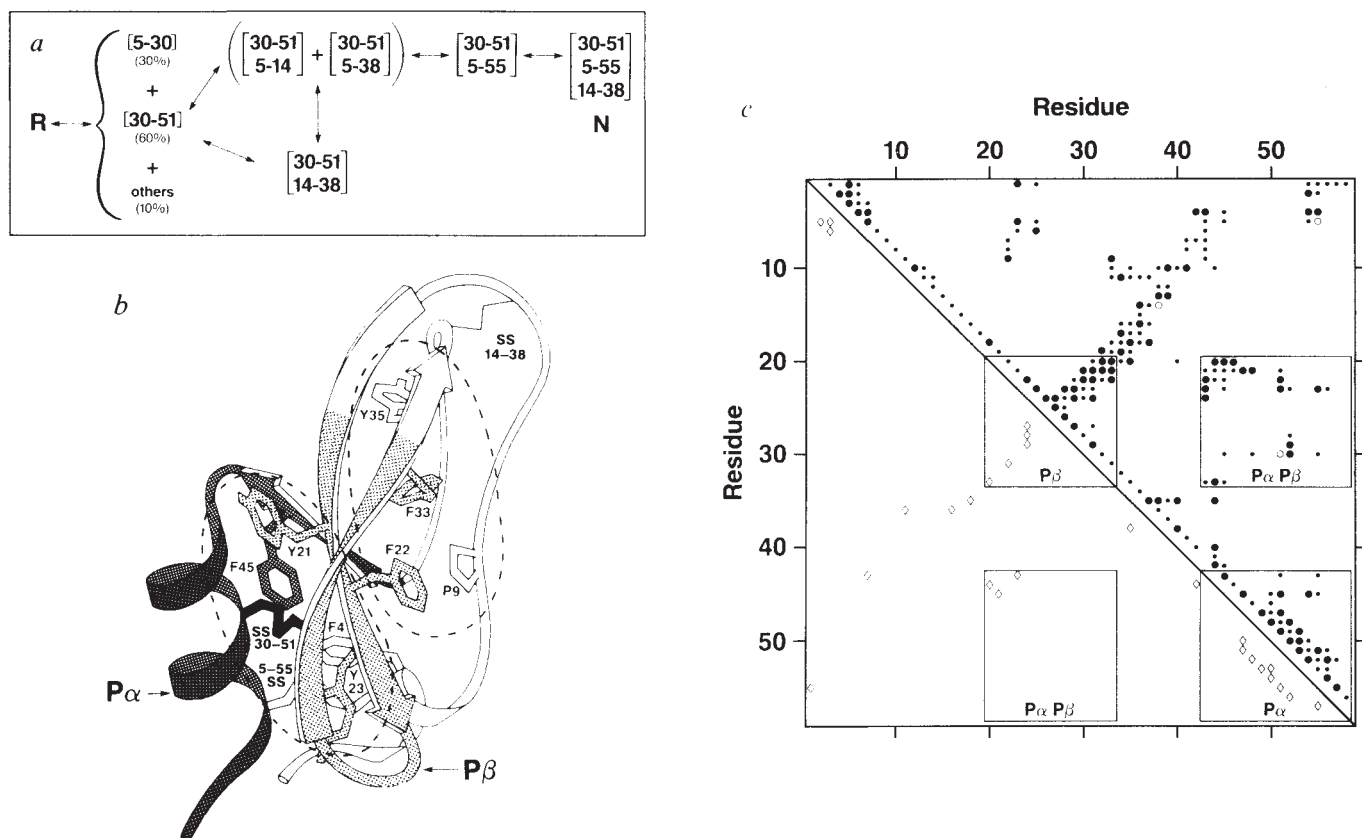


Fig. 1 Design of a peptide model for the [30-51] folding intermediate of BPTI. *a*, Schematic diagram of the folding pathway of BPTI (25 °C, 0.1 M Tris-HCl, 0.2 M KCl, 1 mM EDTA, pH 8.7)³. R designates the reduced, unfolded protein. N refers to the native, folded protein. Only the most populated intermediates are depicted in this diagram, which indicates the disulphide bonds formed at various stages of folding. The non-productive intermediate, N(30SH, 51SH), is not shown. The one-disulphide intermediates are grouped together because they are in rapid equilibrium; the two most populated intermediates are indicated along with their relative proportions. The plus sign between intermediates [30-51, 5-14] and [30-51, 5-38] indicates that they are both formed directly from the one-disulphide intermediates, that they are both converted directly to [30-51, 5-55], and that both are intermediates in the rearrangement of [30-51, 14-38] to [30-51, 5-55]. *b*, Schematic drawing⁵² showing the regions of BPTI corresponding to the peptides $P\alpha$ and $P\beta$. Most of one of the hydrophobic cores in BPTI (indicated with dotted lines) is contained in the region corresponding to $P\alpha P\beta$. *c*, A contact plot of BPTI based on X-ray diffraction data^{17,18}. Van der Waals interactions between any atom of a residue and the atoms of other residues are indicated above the diagonal. If the distance between two atoms is less than 150% of the sum of their van der Waal's radii, a small dot is shown; less than 100% is indicated by a large dot. Radii were used as previously described⁵³. Disulphide bonds are designated with open circles. Intra-residue van der Waal's interactions, and those between residues i and $i+1$ have been omitted for clarity. Hydrogen bonds between residues are indicated below the diagonal.

Methods. The peptides were synthesized on an Applied Biosystems Model 430A peptide synthesizer using standard reaction cycles, modified to include acetic anhydride capping. PAM (4[oxymethyl]phenylacetomidomethyl) resins were used. The peptides were cleaved from the resin using TFMSA (trifluoromethane sulphonic acid) and desalted on a Sephadex G-10 column in 5% acetic acid. The peptide complex ($P\alpha P\beta$) was formed by air oxidation in 5 M guanidine hydrochloride (GuHCl), 0.2 M Tris-HCl, pH 8, 25 °C for 20 h. GuHCl was present to prevent precipitation of the $P\beta$ homodimer. $P\alpha P\beta$ was purified from other products by HPLC on a Vydac C18 semipreparative column using a water/acetonitrile gradient in the presence of 0.1% trifluoroacetic acid. The peptide complex was lyophilized from 5% acetic acid and stored in a desiccator. $P\alpha P\beta$ synthesized by this procedure was indistinguishable, by analytical HPLC, from the product obtained under argon using glutathione as an oxidant. The identity of $P\alpha P\beta$ was confirmed by analysis at the MIT Mass Spectrometry NIH Facility which gave the expected MH^+ relative molecular mass of 3,353.

dominant two-disulphide intermediates formed from [30-51] can easily form a third disulphide bond, even in the case of [30-51, 14-38] which contains two native disulphides³. Instead, an intramolecular disulphide exchange reaction leads to the intermediate [30-51, 5-55]. This last intermediate has a native-like structure and the remaining disulphide bond readily forms between cysteines 14 and 38 (Fig. 1a).

There are two central questions about Creighton's pathway^{3,11}: (1) What is the basis for the accumulation of [30-51] among the one-disulphide intermediates? (2) Why does folding occur preferentially by rearrangements of incorrect two-disulphide intermediates, instead of direct formation of correct disulphides? It is probable that the answer to the first question is structural because [30-51] does not accumulate appreciably when denaturants are present during oxidation¹², and the amount of [30-51] increases both in the presence of stabilizing

Hofmeister salts and at low temperatures¹³. Moreover, hydrodynamic¹⁴, immunochemical¹⁵, ultraviolet absorbance and CD studies¹⁶ indicate that [30-51] has a non-random conformation. An answer to the second question requires an explanation of why Cys 55 does not participate significantly in the second step of the folding pathway (Fig. 1a).

In the design of a synthetic analogue of [30-51], we made the following working assumptions: (1) the structure detected⁷ in [30-51] is localized to regions near the disulphide bond and (2) the structure in [30-51] involves native-like interactions. In BPTI, the 30-51 disulphide connects the C-terminal α -helical region (containing Cys 51) to a strand of the central antiparallel β -sheet (containing Cys 30). The core region of $P\alpha P\beta$ therefore includes residues corresponding to the α -helix and to a substantial part of the central antiparallel β -sheet. Residues outside this core region were also included, using a contact map based

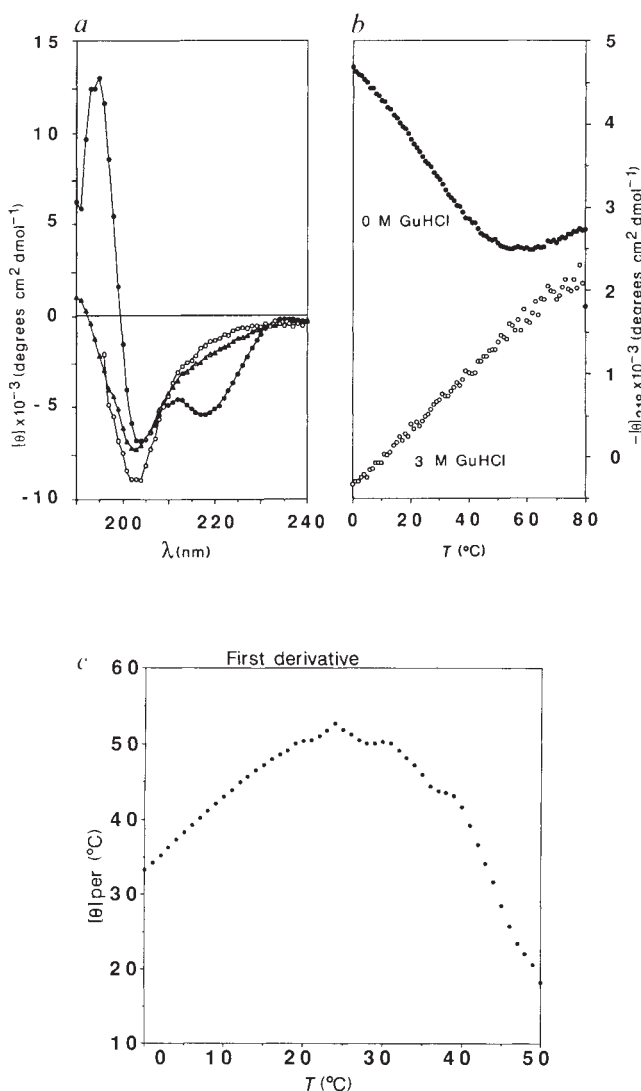


Fig. 2 $P\alpha P\beta$ unfolds in denaturing conditions or when the disulphide bond is reduced. *a*, CD spectra of $P\alpha P\beta$ (disulphide-bonded peptide pair) at 0 °C (●) or 60 °C (Δ); and the reduced complex ($P\alpha + P\beta$) at 0 °C (○). *b*, Temperature dependence of the CD signal at 218 nm for $P\alpha P\beta$, in the presence or absence of GuHCl. *c*, The first derivative of the temperature dependence in the absence of GuHCl. $P\alpha P\beta$ was found to be a monomer under the conditions of these experiments as determined by Sephadex G-25 gel filtration studies. In addition, $[\theta]_{218}$ at 0 °C was independent ($\pm 8\%$) of peptide concentration over the range of 15 μ M to 0.3 mM.

Methods. All spectra were obtained in standard buffer (0.2 M Na_2SO_4 , 10 mM Na_2HPO_4 , pH 6.0) on an Aviv Model 60DS CD spectrometer using a 1-mm pathlength cell at a sample concentration of 0.06 mM (as determined by tyrosine and cystine absorbance at 275.5 nm in 6 M GuHCl; ref. 54). $P\alpha P\beta$ was reduced in 2% mercaptoethanol, 0.5% NH_4HCO_3 and then lyophilized to form reduced $P\alpha + P\beta$, as determined by HPLC. The CD spectrum of reduced $P\alpha + P\beta$ was obtained in the presence of 0.5 mM reduced dithiothreitol and 1 mM EDTA. For the temperature-dependence studies, a 10-mm pathlength cell was used and the sample concentration was 0.02 mM. Temperature was maintained with a HP Model 89100A Peltier temperature control unit. All samples were degassed under vacuum before data collection. A Sephadex G-25 fine column (1.6 cm $\times$ 40 cm) in standard buffer at 4 °C was used for the gel filtration experiments. The calibration standards were, in order of increasing elution volume: reduced, carboxymethylated BPTI; native BPTI; oxidized insulin B chain; Ac-SAEDAMRTAGGA; (PNPA)₃; and Ac-AKFERQHMDs. $P\alpha P\beta$ eluting from the column (at a peak concentration of 0.9 mM) had an apparent relative molecular mass of 2,980, indicating that it was monomeric.

on the crystal structure of BPTI (ref. 17 and 18) as a guide (Fig. 1c). Our aim was to include residues in one peptide that interact with residues in the second peptide (that is, in the native structure) while keeping the peptides as short as possible. $P\alpha P\beta$ represents approximately one-half of BPTI (30 of 58 residues) and it includes most of one of the hydrophobic cores of the protein (Fig. 1b).

The sequences of the individual peptides (called $P\alpha$ and $P\beta$) are:

Asn-Asn-Phe-Lys-Ser-Ala-Glu-Asp-Cys
-Met-Arg-Thr-Ala-Gly-Gly-Ala $P\alpha$: [43-58]

Arg-Tyr-Phe-Tyr-Asn-Ala-Lys-Ala-Gly-
Leu-Cys-Gln-Thr-Phe $P\beta$: [20-33]

$P\alpha$ contains 16 residues and corresponds to residues 43-58, which include the C-terminal α -helix and a short β -strand in BPTI. Cys 55 has been replaced with an alanine, so $P\alpha$ contains only one thiol group (corresponding to Cys 51). $P\beta$ contains 14 residues and corresponds to part of the antiparallel β -sheet in BPTI (residues 20-33) including Cys 30.

The peptides were synthesized chemically using solid-phase methods and the peptide pair was formed by oxidation as described in the Fig. 1 legend. After oxidation, $P\alpha P\beta$ was purified using reverse-phase HPLC. The identity of $P\alpha P\beta$ was confirmed using fast atom bombardment (FAB) mass spectrometry.

$P\alpha P\beta$ folds into a native-like conformation

CD spectra indicate that $P\alpha P\beta$ folds in aqueous solution and that this structure can be unfolded with increasing temperature, whereas there is no evidence for substantial folded structure when the disulphide bond is reduced (Fig. 2a). As monitored by CD, the thermal unfolding transition for $P\alpha P\beta$ is broad, spanning over 60 °C, and a fully folded baseline is not reached even at 0 °C (Fig. 2b). The transition is completely reversible at temperatures up to 80 °C provided that the sample is degassed before use. Adding denaturant (guanidine hydrochloride) eliminates the transition, and a linear temperature-dependence of the CD signal, often seen with unfolded peptides¹⁹ is observed instead (Fig. 2b). $P\alpha P\beta$ does not aggregate, as judged by gel filtration, and there is no significant dependence of the CD signal on peptide concentration over a 20-fold range (see Fig. 2 legend).

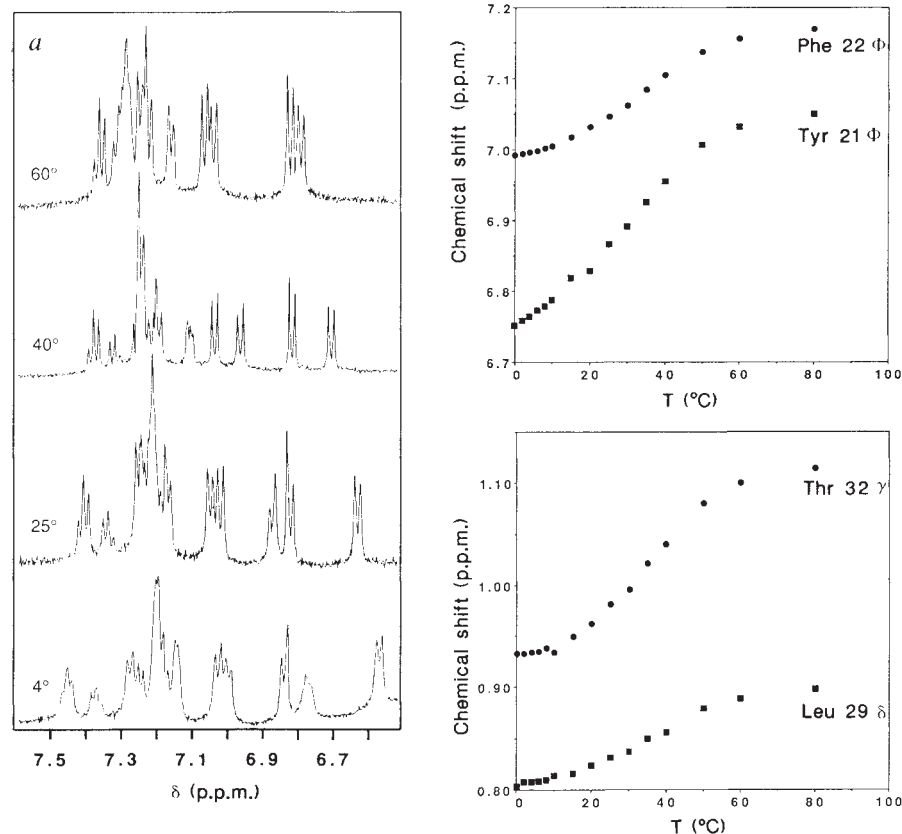
Proton NMR spectra (aromatic region) of $P\alpha P\beta$ at various temperatures are shown in Fig. 3a. There are not separate peaks for folded and unfolded species, indicating that interconversion between folded and unfolded molecules is fast on the NMR timescale. Fast-exchange allows the folding transitions for individual protons to be monitored by chemical shift measurements (Fig. 3b). The temperature-dependence curves are sigmoidal and broad, and different resonances may show qualitatively different transitions, although it is important first to establish proper baseline values for the transition curves.

The folded structure in $P\alpha P\beta$ is similar to that in [30-51] as judged by NMR (compare Fig. 3a with Fig. 3 of ref. 7). The structure in [30-51] seems to be more stable⁷ than that in $P\alpha P\beta$. An interesting difference between the two species is that the interconversion between folded and unfolded molecules of [30-51] shows intermediate-exchange kinetics on the NMR timescale⁷, whereas $P\alpha P\beta$ shows fast-exchange kinetics.

To characterize the structure in $P\alpha P\beta$, we have used two-dimensional NMR spectroscopy. Sequential strategies²⁰ have been used to assign 90% of the resonances in $P\alpha P\beta$. Unequivocal starting assignments were obtained as described in the Fig. 4 legend. A more complete description of the 2D-NMR analysis will be published elsewhere. Figure 4a summarizes experimental data used for backbone resonance assignments. Although the NMR assignments for $P\alpha P\beta$ were made independently of the assignments for BPTI, the relative chemical shifts in the two

Fig. 3 Because the folding of $P\alpha P\beta$ is fast on the NMR timescale, the folding transitions of individual residues can be monitored by chemical shift changes. *a*, Aromatic region of 1H -NMR spectra for $P\alpha P\beta$ at various temperatures. *b*, Temperature dependence of chemical shift changes for selected resonances of $P\alpha P\beta$. Resonances are indicated using nomenclature for the corresponding residue in intact BPTI. p.p.m., Parts per million.

Methods. The sample concentration was 5 mM in standard buffer made with D_2O . The chemical shift standard was TMS (trimethylsilylpropionic acid), whose chemical shift at pH 6.0 was assumed to be -0.017 p.p.m. (ref. 55). The data were collected on a 500 MHz spectrometer at the MIT Francis Bitter National Magnet Laboratory, using pre-saturation of residual HOD for 2.0 s. A shifted sine-bell apodization was used to enhance resolution.



species are remarkably similar (Fig. 4b). For comparison, the range of chemical shifts for amino acid residues in model 'random-coil' peptides²¹ are indicated with squares in Fig. 4b. These results indicate that the structures in $P\alpha P\beta$ and native BPTI are similar and that the population of folded $P\alpha P\beta$ molecules at 4°C is high.

Figure 5a shows a region of a NOESY spectrum of $P\alpha P\beta$. Analysis of the NOESY data indicates that most, if not all, of the secondary structure present in the corresponding region of BPTI is also present in the folded conformation of $P\alpha P\beta$. In particular, the C-terminal α -helix and the central antiparallel β -sheet seem to be intact (Table 1). Tertiary interactions between these regions of secondary structure are also present in $P\alpha P\beta$ (Table 1). These tertiary interactions are similar to those found in intact BPTI, as indicated in Fig. 5b.

Mechanism of protein folding

The finding that $P\alpha P\beta$ folds into a conformation with native-like secondary and tertiary structure supports the initial assumptions made in the design of the analogue (that structure in [30-51] involves native-like interactions and that many of these interactions are localized near the disulphide). Our results indicate that most of the C-terminal α -helix, central antiparallel β -sheet and the hydrophobic core between them is a relatively stable entity in the presence of the 30-51 disulphide. This provides a structural explanation for why [30-51] is populated at high levels in early folding intermediates of BPTI⁶.

The structure seen in [30-51] is thermodynamically coupled to disulphide bond formation. One would like to know if development of secondary and tertiary structure kinetically precedes disulphide bond formation, or vice versa. Creighton's experiments suggest that disulphide bond formation occurs first because all six cysteine residues in BPTI participate equally in the formation of the first disulphide bond²²; rapid disulphide shuffling allows accumulation of the thermodynamically most stable intermediates. Kinetic studies of the association between $P\alpha$ and $P\beta$ may provide a direct evaluation of this conclusion.

Why doesn't Cys 55 participate in the second step of the rearrangement pathway (Fig. 1a)? The structure of $P\alpha P\beta$ could explain why the 14-55 and 38-55 disulphide bonds are not formed readily from [30-51], because the antiparallel β -sheet places Cys 14 and Cys 38 away from Cys 55. The observation that the 5-55 disulphide bond does not form requires another

Table 1 Selected NOEs observed in NMR spectra of $P\alpha P\beta$

Structure	Observed NOESY cross-peak	Distance in crystal structure of BPTI
β -sheet	21 α -32 α	2.4 Å
	23 α -30N	4.4 Å
	24 α -29N	3.4 Å
	26 α -27N	3.5 Å
α -helix	47 β -49N	2.6 Å
	48 β -52N	4.7 Å
	52 α -55 β	4.2 Å
	54N-55N	2.6 Å
Tertiary	21 δ -46 α	4.1 Å
	21 ϵ -48N	3.2 Å
	23 ϵ -55 β	4.1 Å
	23 β -55 β	2.5 Å
	30 β -48 β	3.5 Å
	30N-52 γ	3.4 Å
	45 δ -50 α	4.8 Å
	45 ϵ -54 γ	2.8 Å
	45 ϵ -55 β	4.5 Å

All nuclear Overhauser effects (NOEs) listed were observed as cross-peaks that could be unambiguously assigned (that is, definitive assignments of resonances to the associated protons with no possibility of overlap with other resonances) in NOESY spectra of $P\alpha P\beta$ at 4°C (Fig. 5a). Cross-peaks were observed both in spectra accumulated using 250 ms and 350 ms mixing times. For comparison, the distances between the corresponding protons in the crystal structure of BPTI (ref. 18) are indicated.

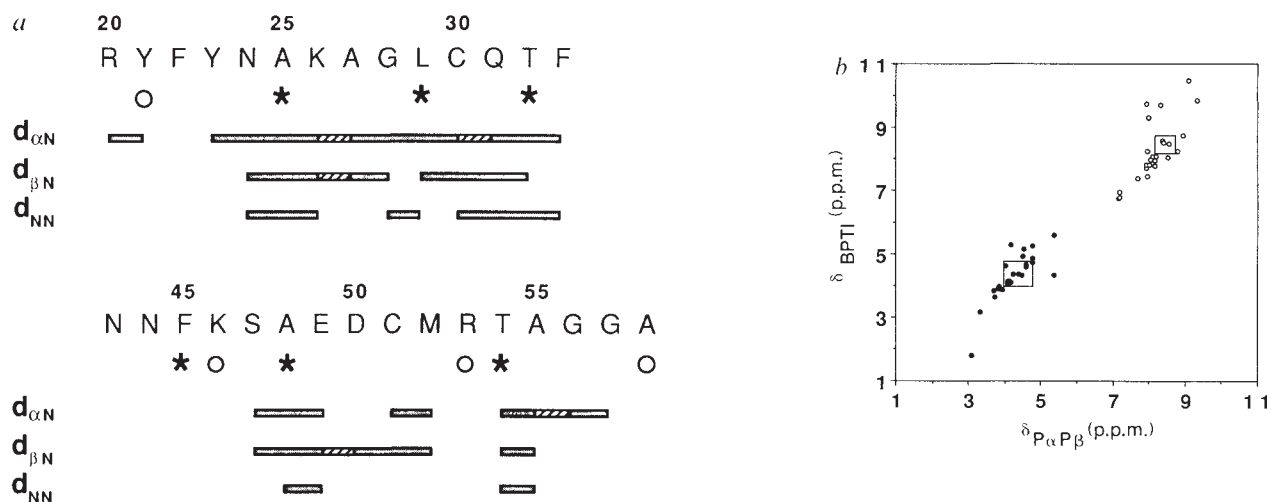


Fig. 4 Independent NMR assignments for $P\alpha P\beta$ give relative chemical shifts that are remarkably similar to those for native BPTI. *a*, Summary of experimental data used for sequential assignments. The amino-acid sequence of $P\alpha P\beta$ (using nomenclature for the corresponding residues in BPTI) is shown and the following symbols are used²⁰: *Definitive assignments made independently from 2D-NMR, as described below; $d_{\alpha N}$, $d_{\beta N}$ and d_{NN} represent NOE connectivities between residues i and $i+1$; dotted regions indicate that an unambiguous NOE was observed; cross-hatched regions that the i to $i+1$ NOE was ambiguous, but that connectivities from this NOE could be unambiguously connected to the appropriate amide and alpha protons on the same residue; O, assignments made on the basis of spin systems after elimination of other possibilities. Essentially complete side-chain assignments have been obtained for all residues except 22, 43 and 44. *b*, The chemical shifts of all assigned amide and alpha protons in $P\alpha P\beta$ against those observed for the corresponding residues in BPTI (ref. 8). The range of chemical shifts values for amide and alpha protons in model 'random-coil' peptides²¹ are indicated with squares.

Methods. DQF-COSY (double quantum filtered two-dimensional J-correlated spectroscopy), RELAY (relayed coherence transfer spectroscopy), TOCSY (total correlation spectroscopy) and NOESY (two-dimensional nuclear Overhauser enhancement spectroscopy)²⁰ spectra (800 $t_1 \times 2,048$ t_2) were collected on a 500 MHz Bruker spectrometer at the Fox Chase Cancer Center (Philadelphia) with a recycle delay of 2.0 s. The mixing times for the NOESY spectra were 250 or 350 ms. Peptide concentration was 15 mM. A sweep width of 5 kHz was used in both dimensions, and a phase-shifted sine-bell function was used to enhance resolution. Definitive assignments that served as starting points for the sequential assignments were made as follows: the methyl resonances of Ala 25 and Ala 48 were assigned using $P\alpha P\beta$ variants which contained deuterated alanine (introduced by peptide synthesis) at those sites. The methyl resonances of the only leucine, Leu 29, were assigned on the basis of the unique COSY pattern. The methyl resonances of Thr 54 and Thr 32, the only threonine residues found in $P\alpha$ and $P\beta$ respectively, were assigned first in spectra of the isolated peptides at high temperatures (60 °C) on the basis of their COSY patterns and intrinsic chemical shift values²¹. These assignments were then transferred to spectra of $P\alpha P\beta$ at 60 °C (that is, unfolding conditions) and the resonances were followed as a function of temperature to 4 °C. From these starting points, sequential assignment strategies²⁰ were used. Details of the assignment procedure will be published elsewhere.

explanation. Previous studies indicate that native-like structure in partially folded intermediates needs to be distorted significantly in the transition state to form the 5-55 disulphide^{23,24}. The native-like structure observed in $P\alpha P\beta$ supports this view, although more work is needed to define the nature of disulphide bond rearrangements in the folding of BPTI.

It has been suggested²⁵ that 'direct' formation of the 5-55 disulphide from [30-51] becomes more significant when folding occurs at higher temperatures. Our results are consistent with this because the structure in $P\alpha P\beta$ and [30-51] (ref. 7) is thermally labile: higher temperatures would be expected to decrease the amount of structure in the vicinity of Cys 55 in [30-51]. The black mamba toxins, which are homologous to BPTI but significantly less stable, follow a more direct folding pathway²⁶. The disulphide rearrangements in the folding pathway of BPTI may be a result of the unusual stability of the protein (see refs 23, 24), although disulphide rearrangements are also observed for less stable proteins such as ribonuclease A (ref. 27) and hen lysozyme²⁸.

In the simple framework model of protein folding²⁹⁻³², marginally stable units of isolated secondary structure form early and then coalesce at a later stage in folding. The interaction between secondary structures is mutually stabilizing. Earlier reports that isolated α -helices and β -turns can form in aqueous solution (see refs 33-35 for examples) provide evidence for the first step of the framework model. CD experiments suggest that a peptide containing the C-terminal α -helix of BPTI can also form an α -helix in isolation (E. M. Goodman and P.S.K., to be published). Our results indicate that interaction of this α -helix with β -sheet regions of BPTI provides a substantial increase in stability, supporting the second step of the framework model.

Moreover, when six residues are removed from the C-terminus of BPTI (that is, including part of the helical region), the amount of [30-51] produced is reduced significantly and the molecule cannot refold³⁶.

Nevertheless, $P\alpha P\beta$ does include most of one of the hydrophobic cores of BPTI (Fig. 1a). Our results do not rule out a folding mechanism in which a non-specific hydrophobic collapse of the polypeptide chain is followed by formation of interacting secondary structure units³⁷, nor have we addressed the possibility of a 'molten globule'^{38,39} intermediate. The roles of specific secondary-structure interactions and non-specific hydrophobic interactions in protein folding remain to be determined.

$P\alpha P\beta$ provides a new model system to study β -sheet formation. It has been difficult to obtain models of β -sheets because intermolecular aggregation often occurs. Where it has been studied in model systems (for example polylysine or polytyrosine), β -sheet formation occurs in the second-to-minute time range^{40,41}. In contrast, the direct folding reactions (that is, uncomplicated by proline isomerization) of several β -sheet-containing proteins, including BPTI (ref. 42), can occur faster than this ($\tau < 50$ ms). Our results indicate that β -sheets can form in less than 1 ms because the interconversion between folded and unfolded $P\alpha P\beta$ molecules is in fast-exchange on the NMR timescale; the frequency difference for these two states can exceed 300 Hz.

Prospects

$P\alpha P\beta$ can be considered as a 'synthetic subdomain' of BPTI, where the term subdomain is used to refer to a unit of folded structure larger than an isolated helix or sheet, but smaller than

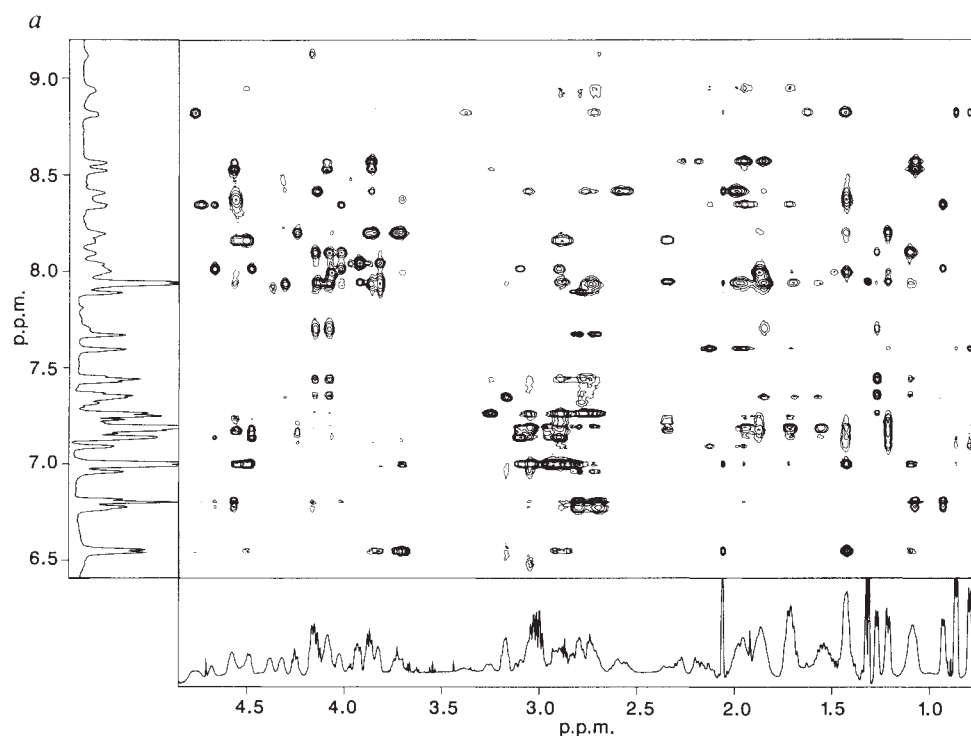


Fig. 5 NOESY spectra indicate that the structure of PαPβ is native-like. *a*, Region of a NOESY spectrum (unsymmetrized) of PαPβ, in 90% H₂O, 10% D₂O, standard buffer, 4 °C. Mixing time was 250 ms. *b*, Stereo drawing of the crystal structure¹⁸ for the region of BPTI corresponding to PαPβ. Selected unambiguous NOEs observed in spectra of PαPβ which result from native-like tertiary interactions are indicated with dotted lines.

an entire domain⁴³. Our demonstration that an isolated subdomain can have substantial and stable native-like structure in aqueous solution supports the validity of hierarchical analyses of protein structure and folding^{43,44}. Our results indicate that protein regions like PαPβ, that are noncontiguous in primary amino acid sequence but which would otherwise correspond to compact units⁴⁵, should be considered in protein structure and folding hierarchies.

The approach used here is best suited for regions of a protein that include a disulphide bond, although in future work it may be possible to connect regions with a non-natural disulphide. Another potential limitation is that the method introduces artificial chain termini. These limitations can be compared with those of other methods for making synthetic subdomains. Proteolytic fragments corresponding to subdomains that fold in aqueous solution have been identified^{46,47}, but typically they are large by standards of normal peptide synthesis. Template-based

synthetic strategies have been developed⁴⁸ that permit studies of the interaction between secondary structure units, but the template is non-native.

The ability to make autonomously folding protein fragments using disulphide-bonded peptide pairs may also have practical applications. Anti-peptide antibodies^{49,50} often have low affinity for native proteins, possibly because the antibodies recognize unfolded conformations. Synthetic subdomains that fold might be useful immunogens. Similarly, synthetic subdomains could be used to map discontinuous epitopes⁵¹ and to design analogues of protein ligands (for example, hormones). Synthetic subdomains from enzymes may prove to be catalytic.

The approach described here for making peptide models of protein folding intermediates and/or subdomains is simple. In particular, large quantities of product can be obtained using standard peptide synthesis methods, and interacting regions that are non-contiguous in primary sequence can be studied without

introducing a non-native linker. By removing a substantial part of the protein, the cooperativity of protein folding is circumvented and structural analysis is simplified. In theory, the free energy of interaction between different protein regions is represented in the equilibrium constant for forming the peptide pair, and the kinetics of association can give information about the mechanism of interaction. The ability to make amino acid changes using peptide synthesis, combined with the ability to carry out detailed NMR analyses will facilitate studies of the structure, stability and folding of peptide models like $\text{P}\alpha\text{P}\beta$.

Received 20 July; accepted 20 September 1988.

1. Roder, H., Elöve, G. A. & Englander, S. W. *Nature* **335**, 700–704 (1988).
2. Udgaonkar, J. B. & Baldwin, R. L. *Nature* **335**, 694–699 (1988).
3. Creighton, T. E. *J. molec. Biol.* **113**, 275–293 (1977).
4. Creighton, T. E. & Goldenberg, D. P. *J. molec. Biol.* **179**, 497–526 (1984).
5. Creighton, T. E. *Prog. Biophys. molec. Biol.* **33**, 231–297 (1978).
6. Creighton, T. E. *J. molec. Biol.* **87**, 603–624 (1974).
7. States, D. J., Creighton, T. E., Dobson, C. M. & Karplus, M. *J. molec. Biol.* **195**, 731–739 (1987).
8. Wagner, G. & Wüthrich, K. *J. molec. Biol.* **155**, 347–366 (1982).
9. States, D. J., Dobson, C. M., Karplus, M. & Creighton, T. E. *J. molec. Biol.* **174**, 411–418 (1984).
10. Goldenberg, D. P. *Biochemistry* **27**, 2481–2489 (1988).
11. Creighton, T. E. in *Protein Folding* (ed. Jaenicke, R.) 427–446 (Elsevier/North-Holland Biomedical, Amsterdam, 1980).
12. Creighton, T. E. *J. molec. Biol.* **113**, 313–328 (1977).
13. Creighton, T. E. *J. molec. Biol.* **144**, 521–550 (1980).
14. Creighton, T. E. *J. molec. Biol.* **87**, 603–624 (1974).
15. Creighton, T. E., Kalef, E. & Arnon, R. *J. molec. Biol.* **123**, 129–147 (1978).
16. Kosen, P. A., Creighton, T. E. & Blout, E. R. *Biochemistry* **22**, 2433–2440 (1983).
17. Deisenhofer, J. & Steigemann, W. *Acta crystallogr.* **B31**, 238–250 (1975).
18. Wlodawer, A., Walter, J., Huber, R. & Sjölin, L. *J. molec. Biol.* **180**, 301–329 (1984).
19. Shoemaker, K. R., Fairman, R., York, E. J., Stewart, J. M. & Baldwin, R. L. in *Proc. 10th Am. Peptide Symp.* (ed. Marshall, G. R.) 15–20 (ESCOM, Leiden, 1988).
20. Wüthrich, K. *NMR of Proteins and Nucleic Acids* (Wiley-Interscience, New York, 1986).
21. Bindi, A. & Wüthrich, K. *Biopolymers* **18**, 285–297 (1979).
22. Creighton, T. E. *J. molec. Biol.* **96**, 777–782 (1975).
23. Creighton, T. E. *J. phys. Chem.* **89**, 2452–2459 (1985).
24. Goldenberg, D. P. & Creighton, T. E. *Biopolymers* **24**, 167–182 (1985).
25. Marks, C. B., Naderi, H., Kosen, P. A., Kuntz, I. D. & Anderson, S. *Science* **235**, 1370–1373 (1987).

We thank R. Rutkowski for assistance with synthesis and purification, S. Almo and J. Fetrow for help with graphics, N. Kallenbach for NMR time, E. Goodman and E. O'Shea for help with NMR analysis, H. Roder and J. Wand for assistance in obtaining NMR data and T. Lin and J. Staley for help with disulphide chemistry. We also thank many colleagues, especially D. Goldenberg, E. Haas, P. Matsudaira and T. Creighton for discussions and T. Alber and C. R. Matthews for comments on the manuscript. T.G.O. is an American Cancer Society Postdoctoral Fellow. This research was supported in part by grants from the NSF and the Lucille P. Markey Charitable Trust.

26. Hollecker, M. & Creighton, T. E. *J. molec. Biol.* **168**, 409–437 (1983).
27. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. Jr *Proc. natn. Acad. Sci. U.S.A.* **47**, 1309–1314 (1961).
28. Saxena, V. P. & Wetlaufer, D. B. *Biochemistry* **9**, 5015–5022 (1970).
29. Pitsyn, O. B. & Rashin, A. A. *Biophys. Chem.* **3**, 1–20 (1975).
30. Karplus, M. & Weaver, D. L. *Nature* **260**, 404–406 (1976).
31. Richmond, T. J. & Richards, F. M. *J. molec. Biol.* **119**, 537–555 (1978).
32. Kim, P. S. & Baldwin, R. L. *A. Rev. Biochem.* **51**, 459–489 (1982).
33. Kim, P. S. & Baldwin, R. L. *Nature* **307**, 329–334 (1984).
34. Dyson, H. J. et al. *Nature* **318**, 480–483 (1985).
35. Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M. & Baldwin, R. L. *Nature* **326**, 563–567 (1987).
36. Creighton, T. E. *J. molec. Biol.* **119**, 507–518 (1978).
37. Rose, G. D. & Roy, S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4643–4647 (1980).
38. Kuwajima, K. *J. molec. Biol.* **114**, 241–258 (1977).
39. Pitsyn, O. B. *J. Protein Chem.* **6**, 273–293 (1987).
40. Hartman, R., Schwane, R. C. & Hermans, J. *J. molec. Biol.* **90**, 415–429 (1974).
41. Auer, H. E. & Patton, E. *Biophys. Chem.* **4**, 15–21 (1976).
42. Jullien, M. & Baldwin, R. L. *J. molec. Biol.* **145**, 265–280 (1981).
43. Rose, G. D. *J. molec. Biol.* **134**, 447–470 (1979).
44. Go, M. & Nosaka, M. *Cold Spring Harb. Symp. quant. Biol.* **52**, 915–924 (1987).
45. Zehfus, M. H. & Rose, G. D. *Biochemistry* **25**, 5759–5765 (1986).
46. Fronticelli-Bucci, C. & Bucci, E. *Biochemistry* **14**, 4451–4458 (1975).
47. Reutimann, H., Luisi, P. L. & Holmgren, A. *Biopolymers* **22**, 107–111 (1983).
48. Mutter, M. in *Proc. 10th Am. Peptide Symp.* (ed. Marshall, G. R.) 349–353 (ESCOM, Leiden, 1988).
49. Lerner, R. A. *Nature* **299**, 592–596 (1982).
50. Walter, G. & Doolittle, R. F. *Genet. Eng.* **5**, 61–91 (1983).
51. Berzofsky, J. A. *Science* **229**, 932–940 (1985).
52. Richardson, J. S. *Meth. Enzym.* **115**, 359–380 (1985).
53. Lee, B. & Richards, F. M. *J. molec. Biol.* **55**, 379–400 (1971).
54. Edelhoch, H. *Biochemistry* **6**, 1948–1954 (1967).
55. DeMarco, A. *J. magn. Reson.* **26**, 527–528 (1977).

LETTERS TO NATURE

Birth of millisecond pulsars in globular clusters

J. E. Grindlay & C. D. Bailyn

Harvard/Smithsonian Center for Astrophysics,
Harvard College Observatory, Cambridge, Massachusetts 02138, USA

The recent report that two millisecond pulsars were found in the globular cluster 47 Tuc¹, together with the previous discoveries in the clusters M28², M4³ and M15⁴, indicates that formation of millisecond pulsars in globular clusters is greatly enhanced over the field, where a fourth example has recently been found⁵. This excess appears to be even larger than that for low-mass X-ray binaries (LMXBs), which show a factor of 100 excess (per unit mass) in globular clusters as opposed to the Galaxy^{6,7}. If the birthrate of millisecond pulsars is indeed greater than that of LMXBs in clusters, then, contrary to the common assumption⁸, LMXBs cannot be the sole precursors of millisecond pulsars. Here we suggest instead that accretion-induced collapse of white dwarfs in binaries⁹ can form millisecond pulsars directly, without requiring a precursor LMXB stage¹⁰. Ablation of the pre-collapse binary companion by the millisecond pulsar's radiation field¹¹, a process invoked to explain some of the characteristics of the recently discovered eclipsing millisecond pulsar^{5,12,13}, can then yield isolated neutron stars without requiring an additional stellar encounter.

A neutron star can be formed by accretion-induced collapse (AIC) of a white dwarf which is driven past its Chandrasekhar

limit by accretion from a binary companion. This can occur for a range of both white dwarf initial masses and of accretion rates¹⁴. An alternative mechanism for the formation of cluster neutron stars is one in which a fraction of the neutron stars formed by the collapse of the cores of massive stars is retained by the shallow potential well of the cluster¹⁵. In view of the large number of millisecond pulsars (MSPs) observed, we suggest that AIC must be occurring at present in globular clusters, because, as we show below, AIC can produce these objects without recourse to a precursor LMXB stage. Ablation of the secondary¹¹ eliminates the necessity for an additional stellar encounter⁹ in the production of isolated neutron stars. It is interesting to note that two of the globular cluster MSPs have long (≥ 100 -day) orbital periods; these are similar to pulsars in the field, for which an AIC origin has been suggested¹⁶.

AIC will result from a C–O or O–Mg–Ne white dwarf of mass $M \geq 1.0$ – $1.2 M_{\odot}$ accreting at a rate of either $\dot{M} < 10^{-9}$ or $\dot{M} > 4 \times 10^{-8} M_{\odot} \text{ yr}^{-1}$. Such massive white dwarfs can arise from the early generation of massive stars (up to $\sim 8 M_{\odot}$; more massive stars which can produce neutron stars directly must be far fewer in number, to avoid disrupting the cluster and polluting it with metals from supernova explosions⁹) or could be produced by the dynamical merging of He white dwarfs¹⁷ from tidal capture or collisions of $0.5 M_{\odot}$ He white dwarfs on cluster giants⁹. Although the low-accretion-rate case may lead to mass loss through nova explosions which inhibit the growth of white dwarf mass¹⁸, the nova mass loss rates are uncertain and in a globular cluster the large number of white dwarf–main sequence star binaries that is expected from tidal capture may statistically favour this possibility. The high-accretion-rate case can be due

to steady accretion from an expanding giant in a >10 -day orbit, and is thus the probable situation for primordial wide binaries in the disk¹⁸. In a globular cluster, such binaries can arise from tidal capture of a main-sequence star into an initially detached configuration, which then overflows its Roche lobe when it leaves the main sequence¹⁹, but they are unlikely to result from tidal capture of giants²⁰. Alternatively, the high-accretion-rate case may arise from dynamical mass-transfer instabilities²¹ when the donor star (in a compact tidal-capture binary) exceeds ~ 0.6 – 0.9 times the mass of the white dwarf primary (depending on the uncertainties in mass loss from the system^{19,21–23}); or it may result from the perturbations in accretion rate induced by passing field stars²⁴ or by captured triple companions²⁵. The close stellar encounters that are required will occur frequently in the dense cores of collapsed clusters, where most LMXBs are found. In such cores, up to 50% of all stars may have tidally captured companions²⁶. It is not entirely clear that close stellar encounters will occur often enough to provide the necessary number of neutron stars in other clusters (such as ω Cen and 47 Tuc²⁷; indeed, no LMXBs, which require an additional tidal capture, are found in these clusters). A detailed quantitative discussion of the exact number of these objects expected in globular clusters is, however, beyond the scope of this paper.

White dwarfs spun up to near-breakup periods of $0.45 R_8^{3/2} M_{1.4}^{1/2}$ seconds (where $M_{1.4}$ is the white dwarf mass scaled by $1.4 M_\odot$ and R_8 is the radius scaled by 10^8 cm) may have radii larger than the $\sim 3 \times 10^7$ -cm values expected for a $1.4 M_\odot$ white dwarf, in fact being closer to 10^8 cm. After collapse, the final neutron star spin period will be $\sim 4(R_8/M_{1.4})^{1/2}$ ms. Thus the spin period of a neutron star produced by AIC is initially in the millisecond range. In the case of high-accretion-rate AIC, it is also likely that the neutron star will be spun up to millisecond periods, because the white dwarf will have accreted either from a massive disk (from disruption of its companion in dynamically unstable mass transfer) or from induced episodes of high accretion rate. In either case, the total spin-up time of the white dwarf is expected to be long enough ($>10^5$ yr) for it to reach its minimum period.

The magnetic field strength is less certain, as white dwarfs in binary systems are known to have a wide range of magnetic field strengths. Post-collapse fields of at least 10^7 G are required for ablation of the secondary to be effective¹¹, but fields in excess of 10^{10} G will cause a rapid spin-down of the neutron star, making it unlikely to become a MSP. Thus the neutron star must be born with a magnetic field smaller than the canonical value of 10^{12} G. The exact value will depend on the field strength of the white dwarf. If we assume flux conservation, a white dwarf radius of 10^7 – 10^8 cm and a neutron star radius of 10^6 cm, the field required for the white dwarf to evolve from AIC into an isolated MSP is in the range 10^5 – 10^6 G. This field strength is significant for another reason. A white dwarf with a strong magnetic field will be spun up to an equilibrium spin period of

$$P = 14 B_7^{6/7} \dot{M}_{-9}^{-5/7} \dot{M}_{-9}^{-3/7} R_8^3 \text{ s} \quad (1)$$

where the magnetic field B_7 and accretion rate $\dot{M}_{-9}$ are scaled by 10^7 gauss and $10^{-9} M_\odot \text{ yr}^{-1}$. Note that in this expression we do not scale the mass transfer rate by the Eddington value, as is usual for neutron stars, so an extra factor of $R^{3/7}$ must be combined with the usual $R^{18/7}$ scaling. If this period is not to exceed the breakup value (at which the equatorial surface velocity reaches the escape velocity), the field must be

$$B_7 > 0.02 M_{1.4}^{1/4} \dot{M}_{-9}^{1/2} R_8^{-7/4} \text{ gauss} \quad (2)$$

For lower values of B , the Alfvén radius is smaller than the physical radius of the star, and equation (1) is invalid. Thus the field strength required for the creation of isolated MSPs by AIC is similar to that at which magnetic effects begin to dominate the accretion process. Such field strengths may be required for collapse to MSPs, because for substantially lower fields collapse may be inhibited by centrifugal effects, whereas for higher fields

the angular velocity may be too low for the collapsed objects to have spin periods in the millisecond range.

A non-magnetic accreting white dwarf will still be spun up to near its breakup period, because that period corresponds to the angular velocity of matter on Keplerian orbits just above its surface. If it is able to collapse, it would form a neutron star with a period of approximately a few milliseconds, but its final magnetic field may be too low for pulsations to be observed. It is possible that high-magnetic-field white dwarfs, such as those presently found in AM-Her stars, could exist in globular clusters and undergo AIC. But in this case, with pre-collapse fields of $\sim 10^8$ G, the resulting pulsars will be 10^{12} -G neutron stars with long (~ 0.1 -s) spin periods, rather than MSPs.

Ablation of the companion requires that it no longer be 'burying' the pulsar with its mass transfer¹¹. This requirement will be generally satisfied for AIC production of neutron stars because collapse from a $1.4 M_\odot$ white dwarf to a $\sim 1.2 M_\odot$ neutron star (with the mass difference lost as binding energy of the neutron star) will widen the orbit and induce eccentricity (as observed in the 47 Tuc pulsars¹). Mass transfer will therefore be inhibited for a low accretion rate, in which situation the companion will no longer fill its Roche lobe, and possibly also for a high accretion rate (but not necessarily for dynamical mass-transfer systems). Thus AIC leads naturally to the low (or non-existent) mass transfer required for ablation of the binary companion.

AIC production of MSPs that have become isolated through ablation of their secondaries may be the mechanism responsible for the production of isolated neutron stars in globulars, which can subsequently be captured onto main sequence stars to yield LMXBs in globulars. This mechanism is much more efficient than recycling the AIC-produced neutron stars out of their parent binary systems by either binary-binary or binary-single star scattering events^{9,23}. Furthermore, the production of isolated neutron stars rather than LMXBs eliminates one objection to AIC¹⁵, namely the large ratio of cataclysmic variables to LMXBs observed in the field ($10^7/50$), which has been proposed as evidence that AIC cannot occur. In our model, one would not expect LMXBs to arise from AIC without subsequent close stellar encounters.

If MSPs are formed by subsequent spin-up of isolated neutron stars, their birth-rates will be similar to LMXBs²³. Isolated MSPs are expected to be detectable for lifetimes comparable to their spin-down times ($\sim 3 \times 10^8$ yr). We note that the suggestion that MSPs must have lifetimes comparable to the Hubble time⁸ is based entirely on the assumption that they evolve from LMXBs. If this were true, the observed large ratio of MSPs to LMXBs indeed requires a very long MSP lifetime, but this argument does not apply to our model, in which MSPs do not evolve from LMXBs. The spin-down time for MSPs are comparable, in turn, to the lifetimes expected for globular cluster burst sources ('bursters')^{23,25,28}, which constitute $\sim 90\%$ of LMXBs in globulars, so the number of MSPs (currently, five have been identified) should be comparable to the number of bursters (nine have been seen) in globulars. But whereas the number of bursters is probably complete²⁸, the surveys for MSPs probably are not—only a few clusters have been surveyed. An indication that the number of MSPs must be much larger than the number of bursters is that two pulsars have been found¹ in one cluster (47 Tuc), whereas the probability of this being the case for bursters is <0.01 (ref. 29). Thus it is highly likely that many more MSPs are produced than bursters in globular clusters. This would be consistent with our proposed AIC origin of pulsars, which does not require a precursor luminous LMXB stage, followed by capture of a fraction of the resulting isolated neutron stars by main sequence stars to form the LMXBs. If a significant number of neutron stars in the field are produced by AIC, this process might be partially responsible for the apparent enhancement of MSPs over LMXBs in the field³⁰.

Our model makes the specific predictions that (1) the number

of MSPs will be enhanced in clusters with central cusps (like M15), where one expects significant tidal capture production of binaries containing massive white dwarfs; and (2) the number of MSPs will be significantly enhanced over the number of cluster LMXBs by a factor approximately equal to the inverse of the probability that an isolated neutron star undergoes tidal capture in the cluster core multiplied by the ratio of the lifetime of a MSP to that of an X-ray burster.

J.E.G. was supported by NASA and the NSF, and C.D.B. by the Harvard Society of Fellows.

Received 14 July; accepted 26 September 1988.

1. Ables, J. *et al.* *IAU Circ.* no. 4602 (1988).
2. Lyne, A. G. *et al.* *Nature* **328**, 399–401 (1987).
3. Lyne, A. G. *et al.* *Nature* **332**, 45–47 (1988).
4. Wolszczan, A. *et al.* *IAU Circ.* no. 4552 (1988).
5. Fruchter, A., Strinebrink, D. & Taylor, J. *Nature* **333**, 237–239 (1988).
6. Katz, J. I. *Nature* **253**, 698–699 (1975).
7. Clark, G. *Astrophys. J.* **199**, L143–L145 (1975).
8. Van den Heuvel, E. P. J. in *Origin and Evolution of Neutron Stars* (eds Helfand, D. & Huang, J.) 393–406 (IAU Symp. 125, 1987).
9. Grindlay, J. in *Origin and Evolution of Neutron Stars* (eds Helfand, D. & Huang, J.) 173–186 (IAU Symp. 125, 1987).
10. Channugam, G. & Becher, K. *Nature* **329**, 696–698 (1987).
11. Ruderman, M. *et al.* submitted to *Astrophys. J.*
12. Kluzniak, W., Ruderman, M., Shaham, J. & Tavani, M. *Nature* **334**, 225–227 (1988).
13. Phinney, E. S., Evans, C. R., Blandford, R. D. & Kulkarni, S. R. *Nature* **333**, 832–834 (1988).
14. Nomoto, K. in *Origin and Evolution of Neutron Stars* (eds Helfand, D. & Huang, J.) 281–304 (IAU Symp. 125, 1988).
15. Verbunt, F. & Hut, P., in *Origin and Evolution of Neutron Stars* (eds Helfand, D. & Huang, J.) 187–198 (IAU Symp. 125, 1987).
16. Taam, R. E. & van den Heuvel, E. P. J. *Astrophys. J.* **305**, 235–245 (1986).
17. Webbink, R. *Astrophys. J.* **277**, 355–360 (1984).
18. Verbunt, F. in *The Physics of Neutron Stars and Black Holes* (ISAS, Tokyo, 1988).
19. Bailyn, C. & Grindlay, J. *Astrophys. J.* **316**, L25–L29 (1987).
20. Bailyn, C. *Nature* **332**, 330–332 (1988).
21. Verbunt, F. & Rappaport, S. *Astrophys. J.* **332**, 193–198 (1988).
22. Grindlay, J. in *Globular Cluster Systems in Galaxies* (eds Grindlay, J. & Philip, A. G. D.) 347–366 (IAU Symp. 126, 1987).
23. Verbunt, F., van den Heuvel, E. P. J., van Paradijs, J. & Rappaport, S. A. *Nature* **329**, 312–314 (1987).
24. Krolik, J., Meiksin, A. & Joss, P. *Astrophys. J.* **282**, 466–480 (1984).
25. Bailyn, C. thesis, Harvard Univ. (1987).
26. Ostriker, J. in *Dynamics of Star Clusters* (eds Goodman, J. & Hut, P.) 347–357 (IAU Symp. 113, 1985).
27. Verbunt, F. & Meylan, G. *Astr. Astrophys.* (to appear).
28. Grindlay, J. in *Evolution of Galactic X-ray Binaries* (eds Trumper, J., Lewin, W. H. G., & Brinkmann, W.) 25 (NATO ASI Ser., 1986).
29. Lightman, A. & Grindlay, J. *Astrophys. J.* **262**, 145–152 (1982).
30. Van den Heuvel, E. P. J., van Paradijs, J. & Taam, R. *Nature* **322**, 153–155 (1986).

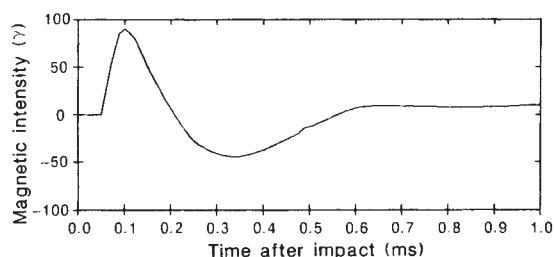


Fig. 1 Vertical component of an inherent magnetic field observed by a coil located just below the impact point. The signal was produced by hypervelocity (5.6 km s^{-1}) impact in a locally reduced ambient field of 500γ .

Transient magnetic fields produced during hypervelocity impact are complex phenomena which depend on the strength of the ambient magnetic field, impact angle, impact velocity and material properties of the target and projectile. Because of the strength of the signals and consistency of the data, we focus on magnetic signals recorded during low-angle (15° from horizontal), high-velocity (4.0 – 6.0 km s^{-1}) impacts by 0.64 -cm aluminum and pyrex projectiles into sand, dry-ice blocks and dolomite powder. The experiments were performed at the NASA-Ames Vertical Gun Range, Moffett Field, California, which is a national impact facility operated jointly by NASA-Ames Research Center and the Lunar and Planetary Institute through NASA's Planetary Geology and Geophysics Program. Typical pressures within the impact chamber were $\sim 1 \text{ mm Hg}$ for impacts into sand and dolomite and $\sim 5 \text{ mm Hg}$ for impacts into dry ice. Detectors were placed below the impact point and/or up to 90 cm down-range from the impact point to ensure that they would be nearby or immersed in the impact-produced plasma.

The magnetic detectors, consisting of several coils in various locations and of varying orientation, provided measurements of dB/dt (where B is the magnetic induction). A centre-tapped coil separated true magnetic signals from electrostatic signals produced by capacitive coupling between the coils and the impact-produced plasma¹¹. A magnetic field with intensity ranging from 0 to $\pm 40 \text{ Oe}$ ($1 \text{ Oe} = 10^3 \gamma$) was applied vertically in the vicinity of the search coils, and data recorded with several oscilloscopes allowed comparison of signal character and arrival times.

With a field of $\pm 2.7 \text{ Oe}$ applied at the impact point, the detected signals show a rarefaction of field strength at the impact point followed by compression and rarefaction as the plasma expands down-range. Following the technique outlined by Lin *et al.*¹¹, we find the conductivity of the cloud from the relation:

$$B = V_0 R \sigma_0 u_1 \Psi_1 \exp(-\alpha)$$

where

$$\Psi_1 = \int_0^1 r' \exp\left(\mu r' - \frac{r'^2 R^2}{b^2}\right) dr'$$

Here B is the integrated rarefaction signal, V_0 is a calibration constant, R is the radius of the expanding plasma cloud, σ_0 is the conductivity at the centre of the cloud, and u_1 is the cloud's expansion velocity; Ψ_1 includes the spatial distribution of the electrical conductivity (μ), the dimensionless spatial variable (r') and the response function of the experimental configuration (a gaussian with half-width b) and α describes the spatial distribution of the cloud's fluid velocity. The system was calibrated by passing an aluminum block at a known velocity over the search coil while embedded in a magnetic field. Evaluating Ψ_1 numerically, the maximum conductivity is found to reach 10^5 – $10^7 \Omega^{-1} \text{ m}^{-1}$ several milliseconds after impact, when the

Laboratory observations of impact-generated magnetic fields

David A. Crawford & Peter H. Schultz

Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA

Radiofrequency emissions have been observed previously during vertical hypervelocity impacts into hard targets, but the dominant magnetic signatures of these events were attributed to amplification of the applied magnetic field^{1–3}. Here we report laboratory experiments which document spontaneous magnetic fields generated by low-angle, hypervelocity impacts in a low-field environment. Low-angle impacts enhance the production of a partially ionized vapour cloud (a dusty plasma) which expands above the impact point^{4,5}. The observed spontaneous magnetic fields may result from runaway thermal electrons producing a toroidal field confined to this plasma and by non-aligned thermal and electron density gradients producing a vertical field within the target^{6,7}. These new observations hold promise both for experimentally probing early-time impact phenomena and for understanding the anomalously high thermal magnetic remanence observed in a young lunar impact melt returned by Apollo 17 (ref. 8). They also may provide clues for understanding the enigmatic record of broad-scale magnetic anomalies on the Moon^{9,10} and anticipated palaeomagnetic signatures from future missions such as the Soviet rendezvous with Phobos.

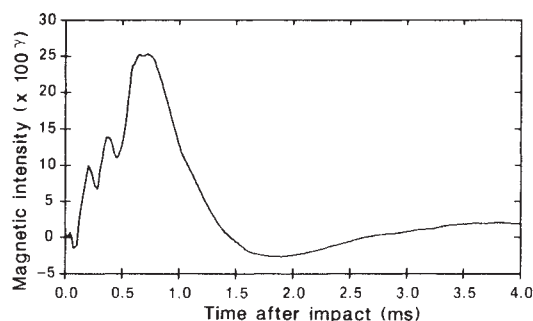


Fig. 2 Horizontal component of an inherent toroidal magnetic field observed by a coil located above the impact point. The signal, produced by runaway thermal electrons in the plasma, attains a maximum intensity significantly greater than the locally reduced ambient field of 500 γ .

Table 1 Magnetic signals from low-velocity impacts

Ambient field* (γ)	Maximum vertical signal* ($\gamma \pm 20\%$)	Maximum toroidal signal ($\gamma \pm 30\%$)	Maximum compressed signal* ($\gamma \pm 25\%$)
$\pm 1.4 \times 10^3$	+36 (2)	320	—
$+6.2 \times 10^4$	+42 (1)	400	—
$+1.4 \times 10^5$	+45 (1)	260	+130
-1.4×10^5	+47 (1)	520	-200
$+2.9 \times 10^5$	+51 (3)	?	+1,400
-2.9×10^5	+40 (2)	?	-1,600

Parentheses indicate number of observations used to determine the result, otherwise only one observation was used.

*Positive is directed upward.

collision frequency between electrons and neutral particles has decreased below a threshold value. The observed conductivity agrees well with predictions for a hypothetical impact on the Moon¹². The high electrical conductivity indicates that the plasma has a significant ionized fraction (10^{-4} to 10^{-3}). Under these conditions, the magnetic Reynolds number, Re_m , and the magnetic diffusion time, τ , are both significantly smaller ($Re_m < 250$, $\tau < 20$ ms) than those expected at some planetary scales; consequently, extrapolation of these measurements should be done with care.

The ambient field was locally reduced to 500 γ by a two-axis cubic array of square (60 cm on a side) field coils, which extended the characterization of the magnetic signals to over three orders of magnitude of ambient field intensity. During high-velocity impacts (5.6 – 5.8 km s⁻¹), a vertical field (Fig. 1) produced by non-aligned thermal and electron density gradients in the impact-induced plasma was observed below and immediately surrounding the impact site within the target. The duration of this field (~ 0.3 ms) and the maximum intensity of 100 γ are identical with the vertical signal observed at much higher ambient field intensities. A toroidal field produced by runaway thermal electrons in the plasma was observed using a horizontal coil above the impact point. It lasted about 1.5 ms with a maximum intensity of 2,500 γ (Fig. 2), which is significantly larger than the reduced ambient field intensity. Table 1 shows the maximum vertical, toroidal and compressed ambient fields observed during ten relatively low-velocity (4 – 5 km s⁻¹) impacts in applied ambient vertical fields of varying direction and intensity.

The above results strongly indicate that spontaneous toroidal and vertical magnetic fields can be produced in a field-free environment by hypervelocity impacts. As summarized in Table

1, the maximum compressed ambient field signature was always less than 1% of the ambient field strength and varied both in sign and magnitude as a function of the ambient field, yet the maximum vertical and toroidal field intensities were roughly constant over two orders of magnitude in ambient field strength. Furthermore, the vertical signal did not change significantly when the applied ambient field was reversed, but the compressed ambient field signatures reversed polarity (Fig. 3).

Impact-generated magnetic fields may explain the thermal magnetic remanence (palaeointensity value 2,500 γ) of lunar sample 70019, an impact melt collected by Apollo 17 from the floor of a 3-m crater⁸. As the melt is only a few million years old, the magnetizing field was probably neither a core dynamo field nor a primordial memory field⁸. Srnka⁷ calculated selected scaling parameters required to determine the extent and intensity of plasma-generated magnetic fields. The magnetic intensity is governed by the amplitude of the temperature and density gradients, which decrease at larger scales, and the timescale, which increases at larger scales. Extrapolating our observations to the scale of the 3-m crater shows that the impact-produced magnetic fields should have intensities comparable to those observed in the laboratory and should last a fraction of a second, long enough for small (~ 1 -mm) droplets of impact melt to cool below the Curie point¹³.

In 1989, a Soviet spacecraft will be able to explore crater-related magnetic anomalies at close range during its rendezvous with the martian satellite Phobos and perhaps will permit application of laboratory results to broader scales. Preliminary experimental observations show that spontaneous magnetic fields produced by hypervelocity impacts have an asymmetric distribution which depends on impact angle. This is qualitatively consistent with the quasisinusoidal magnetic anomalies observed by

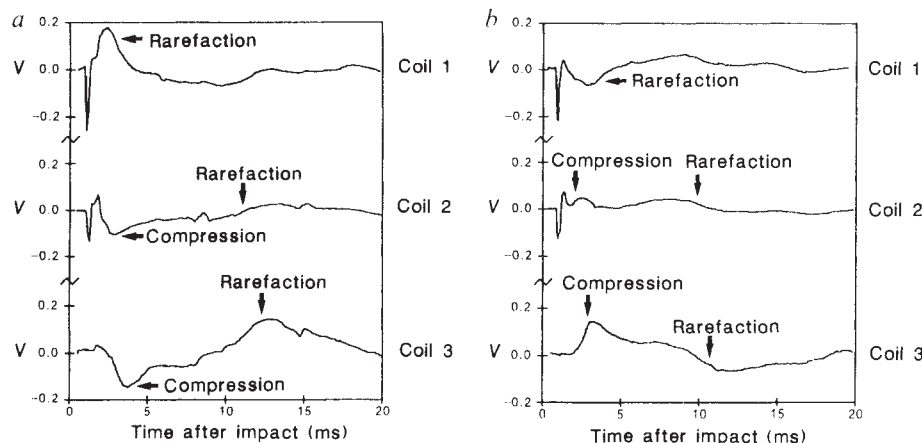


Fig. 3 Signals induced in vertical coils by plasma produced in a hypervelocity (5.6 km s⁻¹) impact. Coil 1 is located below the impact point; coils 2 and 3 are 10 cm and 19 cm down-range, respectively. *a*, The plasma expands in an applied magnetic field of -2.7 Oe (down). The first signal to appear (a down-going pulse on coils 1 and 2) is an inherent vertical field produced by non-aligned thermal and density gradients in the plasma and lasts ~ 0.3 ms. A rarefaction of the applied field is observed below the impact point (coil 1) followed by compression and rarefaction down-range (coils 2 and 3). *b*, The plasma expands in an applied magnetic field of 2.7 Oe (up). The inherent vertical field is observed on coils 1 and 2 exactly as before, but the compression and rarefaction signals change polarity with reversal of the applied field.

Lunokhod-2 as it traversed small (50–400-m) craters on the Moon^{14,15}. With a better understanding of the method of acquiring remanence, the spatial distribution of impact-generated magnetic fields recorded on planetary surfaces (for example, the Moon and possibly Phobos in the future) may provide information on impactor characteristics such as the impact angle, direction and velocity.

We would like to thank Len Srnka for his helpful advice on our experimental technique and for comments on this manuscript.

Received 17 June; accepted 21 September 1988.

1. Martelli, G. & Newton, G. *Nature* **269**, 478–480 (1977).
2. Srnka, L. J. *et al. Earth planet. Sci. Lett.* **42**, 127–137 (1979).
3. Bianchi, R. *et al. Nature* **308**, 830–832 (1984).
4. Schultz, P. H. & Gault, D. E. *Lunar planet. Sci.* **16th**, 740–741 (1985).
5. Schultz, P. H. *Lunar planet. Sci.* **19th**, 1039–1040 (1988).
6. Hide, R. *The Moon* **4**, 39 (1972).
7. Srnka, L. J. *Proc. Lunar Sci. Conf.* **8th**, 785–792 (1977).
8. Sugiyara, N. *et al. Proc. Lunar planet. Sci. Conf.* **10th**, 2189–2197 (1979).
9. Hood, L. L., Coleman, P. J. & Wilhelms, D. E. *Proc. Lunar planet. Sci. Conf.* **10th**, 2235–2257 (1979).
10. Schultz, P. H. & Srnka, L. J. *Nature* **284**, 22–26 (1980).
11. Lin, S., Resler, E. L. & Kantrowitz, A. *J. appl. Phys.* **26**, 95–110 (1955).
12. Hood, L. L. & Vickery, A. *J. geophys. Res. Suppl.* **89**, C211–C223 (1984).
13. Arndt, J., Flad, K. & Feth, M. *Proc. Lunar Sci. Conf.* **10th**, 355–373 (1979).
14. Dolginov, Sh. Sh. *et al. in The Soviet-American Conference on Cosmochemistry of the Moon and Planets* (eds Pomeroy, J. H. & Hubbard, N. J.) NASA SP-370 Part 1, 433–441 (1977).
15. Ivanov, B. A., Okulesky, B. A. & Basilevsky, A. T. in *Impact and Explosion Cratering* 861–867 (Pergamon, New York, 1977).

Limiting mechanisms of slow dilatant plastic shear deformation of granular media

T. Dunstan, J. R. F. Arthur, A. Dalili, O. O. Ogunbekun & R. K. S. Wong*

Department of Civil and Municipal Engineering, University College, London WC1E 6BT, UK

The tendency in granular media to change volume during shear, known as dilatancy, was discussed first in 1886¹ and subsequently by other workers^{2–5}. Strength and resistance to slow deformation can be profoundly affected by dilatancy and this is important for the bulk handling of materials in the process industries and soil behaviour in civil engineering. Here we propose a minimal form of local collapse as a mechanism of slow steady-state deformation, complementing the familiar mechanism of potentially expansive sliding and rolling. Both are applied to ideal frictionless grains and involve, respectively, maximum and minimum rates of internal energy dissipation. Subsequent comparison to the behaviour of real granular material shows surprising similarities. Apparently, our proposed very simple idealization identifies the mechanisms in real materials and provides the possibility of quantitative predictions.

Steady state implies the sustaining of a given principal stress ratio σ_1/σ_3 and dilation rate $-\delta\epsilon_3/\delta\epsilon_1$ through a finite strain increment (see Fig. 1a, which defines the stresses σ and deformations $\delta\epsilon$). Both the conventional mechanism and our proposed local-collapse mechanism apply to generalized stress and strain states, but here we present data from plane strain tests on sand which conform to the mechanisms for the ideal material, with a simple amendment accounting for frictional grains.

Two-dimensional photoelastic⁶ and computer models⁷ of shearing granular materials have consistently shown concentrated contact force lines through the material in the general direction of the major principal stress. Although these force lines appear rather stable during strain for the two-dimensional case, this stability seems unlikely to be maintained in real

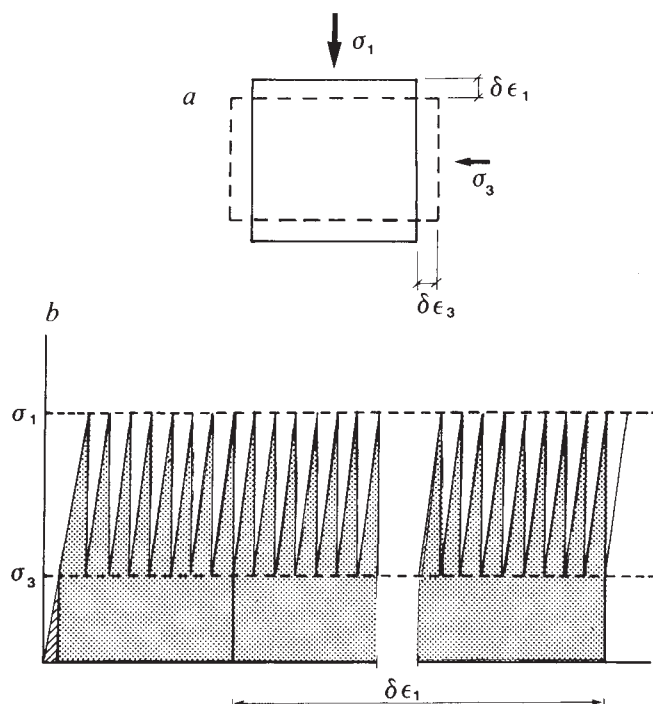


Fig. 1 A conceptual representation of the energy dissipated from collapsing columns during a strain increment $\delta\epsilon_1$ while supporting a major principal stress σ_1 . a, Geometry of applied stresses σ_1 and σ_3 , and strain increments $\delta\epsilon_1$ and $\delta\epsilon_3$. b, Stress-strain relationship, showing the elastic energy stored on application of confining stress σ_3 . The hatched area is the energy stored on initial application of σ_3 , and the stippled area corresponds to energy stored and dissipated in a transitory fashion.

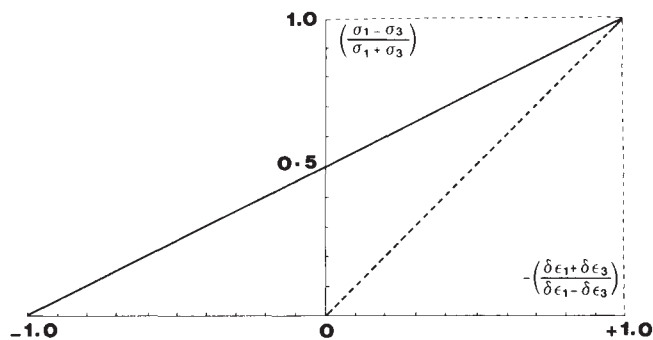


Fig. 2 The new stress-dilatancy relationship (solid line) and Cole's modified stress-dilatancy relationship⁸ (dashed line) for perfectly smooth granular media.

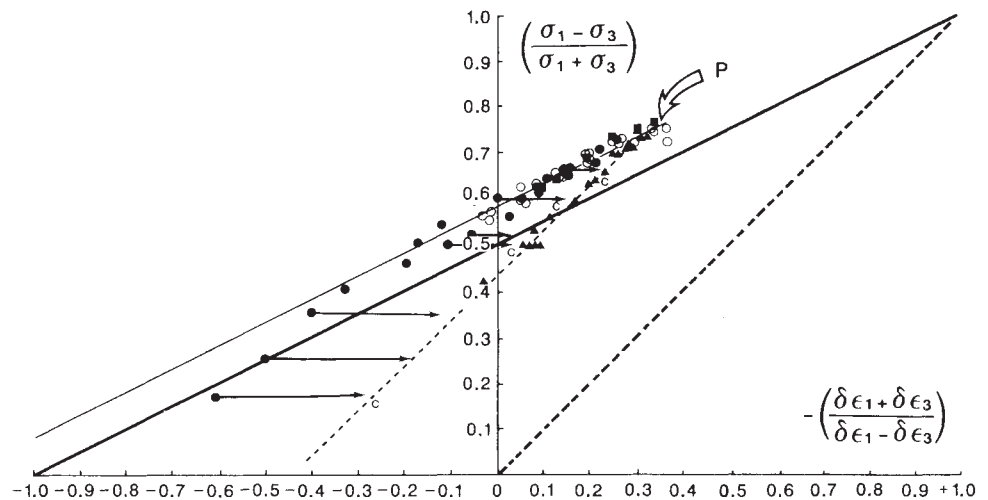
three-dimensional packings. Accordingly, granular material may be modelled as a changing collection of spatially dispersed crooked columns which are supported by distributed lateral forces arising from the minor principal stress and which transmit all the major principal stress through the material. During the very brief existence of each of these transitory columns in an infinitesimal strain increment, the column stores elastic energy, just as would a solid column, and on collapse the stored energy is released and dissipated. As this process is continuous throughout the finite strain increment, in that there are always energy-storing columns, the quantity of energy, E , that is momentarily stored and then dissipated is the same as that which would have been stored in a continuous compressive linear elastic spring which had been initially loaded to the σ_3 level:

$$E = (\sigma_1 - \sigma_3)\delta\epsilon_1/2 + \sigma_3\delta\epsilon_1$$

Figure 1b illustrates the incremental process. The small amount of elastic energy that is stored cumulatively from when the

* Present address: Ove Arup and Partners, London, UK.

Fig. 3 Data from plane strain tests superimposed on the theoretical lines from Fig. 2. ○—Cambridge simple shear apparatus; ●—University College, London (UCL) directional shear cell (cyclic); ■—UCL biaxial tester; ▲—UCL directional shear cell. The latter two use constant principal stress directions. Thin solid and dashed lines denote empirical upper and lower limiting lines, respectively.



confining stress σ_3 is first applied is shown hatched, and is negligible in relation to the energy stored and dissipated in a transitory fashion, which is shown stippled. The form of collapse is limitingly slow, as it must be if, for constant principal stress directions, the granular material will support a stress ratio σ_1/σ_3 just less than that currently applied, without straining further. There is then no direct energy dissipation through collapse, so this elastic storage and release will be the only way in which energy can be dissipated. Under this restriction, the rate of energy dissipation is maximal. Now specializing to plane strain ($\delta\epsilon_2 = 0$), the incremental work-internal energy dissipation equation is

$$\sigma_1 \delta\epsilon_1 + \sigma_3 \delta\epsilon_3 = (\sigma_1 - \sigma_3) \delta\epsilon_1 / 2 + \sigma_3 \delta\epsilon_1$$

which simplifies to

$$-\sigma_3 \delta\epsilon_3 = (\sigma_1 - \sigma_3) \delta\epsilon_1 / 2$$

This is the new stress-dilatancy equation for the slow-collapse mechanism of frictionless granules; it is plotted in the form which is conventional in soil mechanics in Fig. 2. The stress-dilatancy equation for the same granules and zero internal energy dissipation is shown as a dashed line in Fig. 2 and corresponds to Cole's relationship⁸ when the shear resistance at constant volume is set to zero, as appropriate for perfectly smooth grains. The mechanism is one of sliding and rolling with no collapse. The intersection of the two lines represents the limit at which $\delta\epsilon_1 \rightarrow 0$ and can be considered to represent the behaviour of an assembly of frictionless blocks with faces set normal to the major principal stress direction. The line in Fig. 2 corresponding to the new relationship intersects the horizontal axis as $\delta\epsilon_3 \rightarrow 0$ (this represents one-dimensional consolidation, familiar in soil mechanics).

The implicit assumptions of linear contact force compliance and of isotropy of the ideal material are unlikely to apply for real granular materials. Unloading columns to a fixed level, equivalent to the minor principal stress, also seems improbable, but data from tests on sand, in which 90° jumps in the principal stress directions were imposed, show a response that is likely to maintain this loading⁹. The absence of complete unloading may make the assumption of linearity a sufficiently good approximation. Real frictional granular materials may shear with both collapse and sliding/rolling mechanisms operating simultaneously. This will lead to a range of possible dilatancies for a given stress ratio and may be expected to incorporate any anisotropy observed in the stress-dilatancy relationship.

Figure 3 superimposes on Fig. 2 data from plane strain shear tests on Leighton Buzzard sand. Data from three different sets of experiments are distinguished. Cyclic directional shear cell tests, in which the principal stress directions were cycled while a constant stress ratio was maintained, provided a unique range

of unusually accurate stress-dilatancy data, because constant dilatancy rates were maintained throughout large strains¹⁰. In some cyclic tests dilatancy rates progressively increased until a constant rate was achieved. In these tests the rate of accumulation of shear strain per cycle decreased most markedly and in some cases, shear strain disappeared. For these tests, in Fig. 3, the initial dilatancy is connected to the final dilatancy by an arrow. A letter C by the arrow-head indicates that shear strain ceased. The top two lines of this kind in Fig. 3, which are marked with a C in this manner, are taken from different cyclic tests in which only the mean stress level was cycled while principal stress directions were constant. Strain increments in these tests were measured for complete cycles and a steady state was indicated by constant dilatancy. Dilatancy rates within a steady-state cycle undoubtedly vary when the principal stress directions rotate. Paradoxically, strain generated when the principal stress direction is at the centre of its range, with coincident principal strain rate, appears to control the whole cycle. The limitations imposed by coincident axes of stress and strain rate, which are inherent in the model, are thus met in an unexpected way.

The data show limiting stress-dilatancy relationships which are linear and are parallel to the ideal ones. Many of the data points⁹⁻¹¹ describe a line parallel to the collapse stress-dilatancy relationship and displaced only slightly from it. Moreover, all material failure points, defined by the maximum attained shear stress ratio $(\sigma_1 - \sigma_3)/(\sigma_1 + \sigma_3)$, fall on the upper limiting line. That the displacement from the ideal relationship is small implies that dissipation through friction is small¹², and indeed it may be that the discrepancy is in fact due to variations in local constraint on stability. The mechanism at failure is that of virtually pure slow collapse. Published data suggest that this mechanism holds for stiff irregularly shaped particles, which would include all soils with not more than, say, 20% clay fraction. This interpretation may be challenged, because Rowe's stress-dilatancy theory¹³ shows a good fit to the data parallel to the upper theoretical line in Fig. 3 (provided the negative dilatancies are excluded). But even this fit depends on the assumption that the intercept on the vertical axis at zero dilatancy is due to internal friction, whereupon this point provides a reference for fitting the theoretical curve to the data.

The internal dissipation of energy by sand-grain sliding is probably the cause of the much larger offset of the amended lower limiting line. But at present this is an empirical observation based on tests on one material, and further work is needed. Point P on Fig. 3 appears to represent the upper limit of operation of the collapse mechanism.

Limited data¹⁴ suggest that soft spherical particles of materials such as polypropylene fail in a combined mechanism that is dominated by rolling and sliding. However, irregular blocky particles of Teflon, which have very low surface friction, come

quite close to the collapse limit, demonstrating that it is generally important. A current series of monotonic and cyclic tests on damp coal indicates behaviour similar to that of sand.

Stress-dilatancy mechanisms are of potential importance to diverse fields. In steel making, as in any process industry, the reliable provision of material constituents is essential to economic production. Here mass-flow coal hoppers are often duplicated in an attempt to avoid unexpected failures in flow. A proper understanding of stress dilatancy will assist design for reliable flow. The remarkable stiff sand found on the bed of some parts of the North Sea may be explained by the stress-dilatancy response to cyclic loading by the waves. A better understanding will improve estimates of safety margins for offshore oil installations.

We thank Mr F. Cutler, Mr J. Ford and Mr J. Pulsford for making the equipment.

Received 13 June; accepted 27 September 1988.

1. Reynolds, P. *Proc. R. Inst.* **2**, 354–363 (1886).
2. Taylor, D. W. *Fundamentals of soil mechanics* 354–359 (Wiley, New York, 1948).
3. Rowe, P. W. *Proc. R. Soc. A* **269**, 500–527 (1962).
4. Bishop, A. W. *Geotechnique* **16**, 89–130 (1966).
5. Papadopoulos, J. M. thesis, Massachusetts Institute of Technology (1986).
6. Drescher, A. & de Josselin de Jong, G. *J. Mech. Phys. Solids* **20**, 337–351 (1972).
7. Cundall, P. A., Drescher, A. & Strack, O. D. L. *Proc. IUTAM Conf. Delft* 355–370 (1982).
8. Cole, E. R. thesis, Cambridge Univ. (1967).
9. Wong, R. K. S. & Arthur, J. R. F. *Geotechnique* **35**, 471–481 (1985).
10. Wong, R. K. S. & Arthur, J. R. F. *Geotechnique* **36**, 215–226 (1986).
11. Stroud, M. A. thesis, Cambridge Univ. (1971).
12. Skinner, A. E. *Geotechnique* **19**, 150–157 (1969).
13. Rowe, P. W. *Geotechnique* **19**, 75–86 (1969).
14. Sharma, Y. K. thesis, London Univ. (1976).

Acoustic and microwave signatures of breaking waves

W. K. Melville, Mark R. Loewen, Francis C. Felizardo, Andrew T. Jessup & M. J. Buckingham*

R. M. Parsons Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

* Radio & Navigation Department, Royal Aircraft Establishment, Farnborough, Hampshire GU14 6TD, UK

Field experiments have shown that breaking ocean waves may register as discrete events in passive acoustic and active microwave remote sensing measurements of the ocean surface. Using a laboratory model of deep-water wave breaking^{1,2} we show here that the noise generated by breaking and the radar cross-section of breaking waves both correlate with the dissipation of surface wave energy. Our results imply that these remote sensing techniques ultimately may be used to measure the dynamics of breaking waves, and are not restricted simply to obtaining the statistics and kinematics of breaking.

It has long been recognized that breaking is an important mechanism for the dissipation of ocean waves, but only recently has its larger role in the dynamics of the upper ocean begun to be recognized and quantified. Until recently, progress has been slow because of the difficulty of measuring spatially and temporally intermittent events in what is often a hostile environment. These difficulties have led a number of investigators to seek other means of measuring breaking, using microwave or acoustic techniques. Thorpe³ has pioneered the use of active acoustics to image the bubble clouds generated by breaking and to use them as tracers of the boundary layer. Farmer and Vagle⁴ have used passive acoustics to identify breaking events as one of the major sources of surface noise in the ocean, and have shown that the acoustics contain information about the wave field. Phillips⁵ has suggested that a significant component of microwave scattering from the ocean surface may be due to short

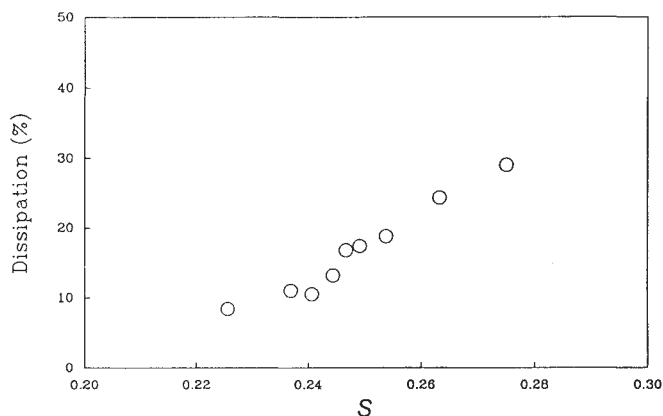


Fig. 1 Fractional dissipation of the energy in the wave packet as a function of the slope parameter S . The onset of breaking occurs at $S=0.241$, and losses at lower values of S are due to other dissipation mechanisms in the channel (for example, viscous dissipation in the boundary layers). Note that to lowest order the fractional dissipation is equal to the fractional loss of excess momentum flux¹. Each data point is the average of at least two repeats of the experiment.

episodes of high scattering cross-section ('sea spikes') associated with breaking waves. It remains, nevertheless, to establish whether there are quantitative correlations between the dynamic and kinematic variables that describe breaking and the microwave and acoustic signatures, that is, whether breaking can be quantified, rather than merely detected, by these remote sensing techniques. Here we present evidence to show that this is the case.

Recently we have shown^{1,2} that controlled experiments on breaking can be conducted in the laboratory and that the dissipation, turbulence and mixing resulting from breaking can be correlated with a few dimensionless variables that characterize the pre-breaking wave field. Of greatest importance is the fact that the dissipation, or momentum flux from waves to currents, depends most strongly on a dimensionless amplitude parameter, S , that describes the wave spectrum. Here $S = Na_i k_i$, where N is the number of spectral components, a_i is the amplitude of the component and $k_i = f_i^2/g$, with f_i the frequency of the component. In these experiments the slope $a_i k_i$ of each component is held fixed, and we measure the radiated noise and microwave scattering from generated breaking waves.

Breaking waves were generated in a steel-framed glass channel 25 m long, 38 cm wide and 38 cm deep. A piston wave generator was programmed to focus a dispersive wave packet at a point 8 m down the channel. The wave packet was synthesized from 32 components of constant slope ak over a bandwidth of 0.64 Hz, spanning the range 0.56–1.2 Hz. The strength of breaking was controlled by varying the slope while keeping other parameters the same. The surface wave field was measured with an array of resistance wave gauges upstream and downstream of breaking, from which the dissipation resulting from breaking was determined as a function of the slope S (Fig. 1).

An X-band cw Doppler radar with a 15-cm aperture horn was mounted looking upstream (towards the oncoming waves). The wave channel was fitted with microwave-absorbing material to ensure that the only significant backscattering came from the surface waves. Preliminary experiments showed that at large angles of incidence multiple scattering was significant. To avoid this, we operated the radar at an incidence angle of 65°, with the aperture of the horn 71 cm above the quiescent water surface. The 3-db footprint of the radar extended 1 m along the water surface.

The sound generated by breaking was monitored with a B&K Model 8105 hydrophone mounted on an L-shaped support and pointing towards the breaking region. The hydrophone was kept

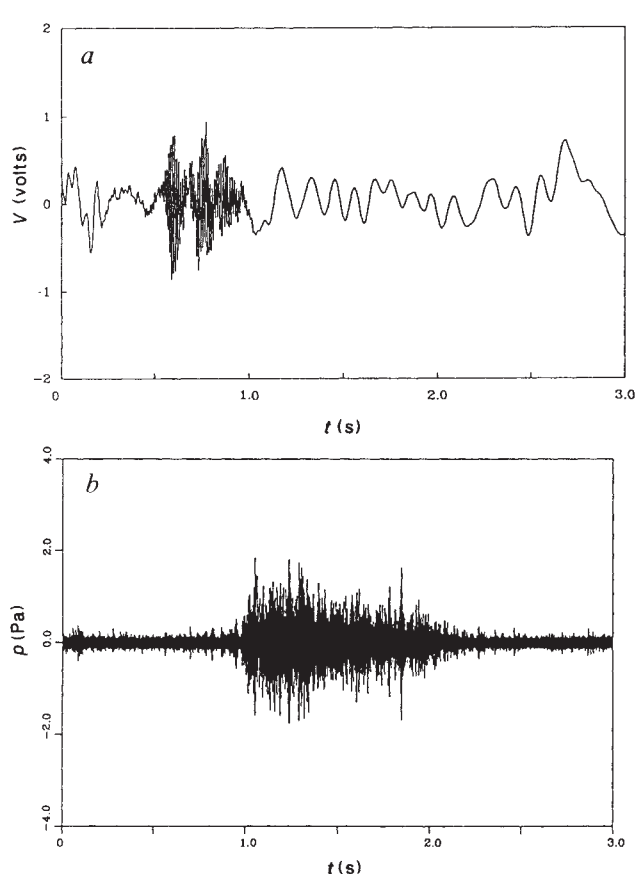


Fig. 2 Time series of the raw data measured for a typical breaker plunging at 7.5 m; $S=0.254$. *a*, Radar backscatter measured at 8.5 m; *b*, Hydrophone signal measured at 15.5 m. Note the higher-frequency components which distinguish the breaking event in the radar signal. In subsequent figures the radar data has been band-passed in the range 50–250 Hz before additional processing.

at mid-depth to maximize the response to the lower acoustic modes. Measurements were made at a number of stations along the tank to account for attenuation along the channel. Surface gravity waves and acoustic waves were absorbed by a 'hog's hair' beach at the end of the channel. Tests showed that the reverberation time of the channel was ~ 100 ms, which is one order of magnitude less than the duration of the acoustic events that we measured. Preliminary measurements showed that the noise generated by breaking was divided between contributions at lower frequencies, up to 1,000 Hz, and those above the cutoff frequency at $\sim 2,600$ Hz. The lower-frequency components were contaminated by noise from the hydraulic wavemaker, so the measurements reported here were high-pass filtered at 500 Hz before digital sampling at 20 kHz. The wave-gauge signals were sampled at 100 Hz and the radar signal at 500 Hz.

For low values of S , breaking did not occur. On slowly increasing the slope a visual breaking threshold was obtained at $S=0.241$. It was found that the visual and acoustic thresholds for breaking were coincident—that is, no acoustic event was observed below this visual threshold. But because capillary waves were generated on the forward face of steep (but unbroken) gravity waves, the radar scattering cross-section increased above the background noise level at values of S below the visual breaking threshold.

Figure 2 shows an example of the signals measured by the radar and the hydrophone for a single plunging event. The duration of each acoustic event was found to be within 15% of the duration of visible breaking at the surface. The duration of the breaking (high-frequency) event in the radar signal is shorter than that of the acoustic event but this is due in part to the fact

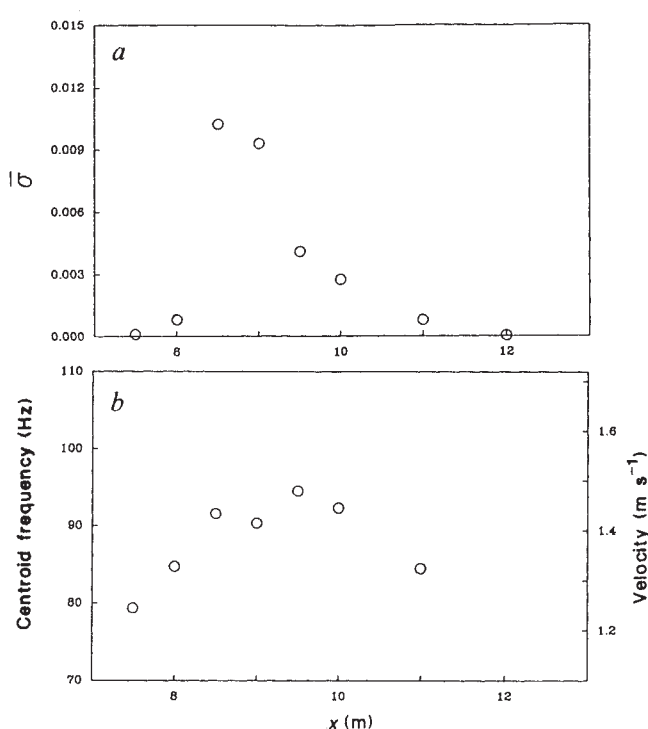


Fig. 3 *a*, Time-averaged radar backscattering cross-section, $\bar{\sigma}$ (in arbitrary units), for a plunging breaker ($S=0.254$) measured at different locations along the tank. Each realization was averaged over 3 s. Three measurements were made at each position and ensemble averaged. *b*, Centroid frequency and corresponding horizontal velocity of the averaged Doppler spectrum for the measurements in *a*. Note that the determination of the centroid at $x=12$ m was not meaningful as the band-passed signal was comprised almost entirely of 60-Hz noise and its harmonics.

that the footprint of the radar is shorter than the maximum length of the surface directly affected by breaking. The fact that the concave surface of the wave (immediately before breaking) is a good reflector may also contribute to the donation of the microwave event. To avoid difficulties associated with accounting for the antenna pattern of the radar and to arrive at a single measure of the backscatter from a breaking event, the radar measurements were taken at a number of positions along the channel. Spectral analysis of the radar data at values of S for which breaking occurred showed that the higher-frequency components associated with breaking could be isolated with a band-pass filter in the range 50–250 Hz. The 50-Hz lower limit offered a clear separation from the lower frequencies, which were associated with the orbital velocity components. Figure 3 shows an example of the measured radar cross-section as a function of position for a single breaking event, along with the centroid of the Doppler spectrum. The centroid is a measure of the mean velocity of the scatterers. In this case the centroid frequencies correspond to velocities in the range of the phase velocities of the components comprising the packet, 1.3 – 1.8 m s $^{-1}$. The cross-section is negligible outside the breaking region, so an unambiguous average backscattering cross-section may be measured.

To account for the attenuation of the acoustic modes along the channel, measurements were taken at a number of stations downstream of the breaking event. The mean-square pressure, p^2 , decayed at an approximately constant rate along the channel, sensibly independent of S . For reasons of brevity we have chosen to display p^2 measured at 15.5 m, approximately four characteristic wavelengths downstream of the breakpoint. We have also integrated the time-averaged radar cross-section, $\bar{\sigma}$, along the channel to give a single space-time-averaged value, $\langle \bar{\sigma} \rangle$, for each event. Both p^2 and $\langle \bar{\sigma} \rangle$ are plotted as a function of S in Fig. 4a. These two variables are very strongly correlated and

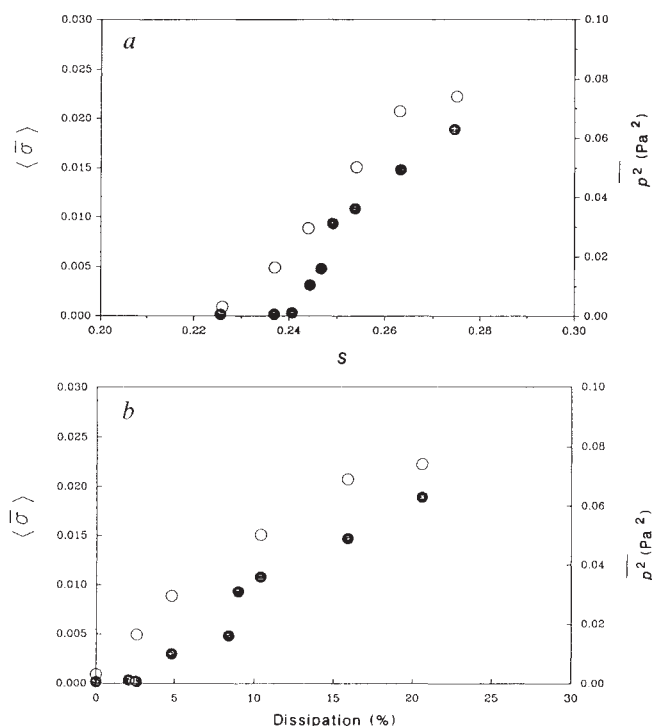


Fig. 4 *a*, The mean-square pressure, $\overline{p^2}$ (averaged over 4 s), (●) and the spatial average of the time-averaged radar cross-section, $\langle \bar{\sigma} \rangle$ (○), both plotted as a function of the slope parameter S . $\langle \bar{\sigma} \rangle = \int_{x_1}^{x_2} \bar{\sigma} dx$, where $\bar{\sigma}$ is given by results similar to Fig. 3a. *b*, p^2 and $\langle \bar{\sigma} \rangle$ (symbols as in *a*) re-plotted as functions of the fractional dissipation, using data from Figs 4a and 1.

both increase monotonically with S . The data from Figs 1 and 4a are combined in Fig. 4b to show the correlation of p^2 and $\langle \bar{\sigma} \rangle$ with the fractional dissipation (or loss of excess momentum flux through breaking)^{1,2}. These data show that, within the scatter of the measurements, p^2 and $\langle \bar{\sigma} \rangle$ vary linearly with the dissipation that results from breaking.

These results show that both the noise generated by breaking and the radar cross-section of breaking waves can be correlated with important dynamical variables. In particular, if these results are shown to extend to the ocean, we anticipate that such measurements can be used to infer the dissipation and mixing from breaking at the ocean surface. For example, the data presented here would support the hypothesis that the radiated acoustic energy is proportional to the dissipated wave energy. Wilson⁶ has proposed that the ambient noise in the ocean is proportional to the whitecap coverage, which is, in turn, approximately proportional to u_*^3 , the cube of the friction velocity^{7,8}. This would appear to be consistent with our measurements, which show that the acoustic energy radiated is approximately proportional to the gravity wave dissipation, which is also believed to scale with u_*^3 . More precisely, we have found that the acoustic energy radiated is proportional to the duration and length of the breaking event which is the laboratory analogue of whitecap coverage.

We thank our colleagues in Electrical Engineering, MIT, MIT Lincoln Laboratory, and Bill Keller at NRL for advice and the loan of equipment. We thank Jack Croker for the design and construction of equipment, and Dr Tom Spence (formerly of ONR) for his support and encouragement during the initial stages of this research. This work was supported by an MIT/Sloan Grant, an ONR (Physical Oceanography) contract and an NSF grant to W.K.M. The participation of M.J.B. was supported by ONR (Arctic Programs).

Received 29 July; accepted 2 September 1988.

1. Melville, W. K. & Rapp, R. J. *Nature* **317**, 514–516 (1985).
2. Rapp, R. J. & Melville, W. K. *Phil. Trans. R. Soc. Lond.* (submitted).

3. Thorpe, S. A. *Phil. Trans. R. Soc. Lond.* **A304**, 155–210 (1982).
4. Farmer, D. M. & Vagle, S. J. *geophys. Res.* **93**, 3591–3600 (1988).
5. Phillips, O. M. J. *geophys. Res.* (in the press).
6. Wilson, J. H. J. *acoust. Soc. Am.* **63**, 952–956 (1980).
7. Monahan, E. C. & O'Muirheartaigh, I. G. J. *remote Sensing* **7**, 627–642 (1986).
8. Wu, J. J. *phys. Oceanogr.* **9**, 1064–1068 (1979).

Evidence from Fram Strait (78° N) for early deglaciation

Glenn A. Jones & Lloyd D. Keigwin

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

Recent syntheses of the history of the last Northern Hemisphere glaciation and deglaciation illustrate that understanding of the mechanisms and timing of deglaciation before approximately 12,000 years BP^{1–5} is limited. After 12,000 yr BP, however, there is sufficient evidence from radiocarbon-dated moraines, raised beaches, varved lake sediments and pollen records to provide a reasonable temporal and geographic picture of the decay of the ice sheets. Here we report on the first oxygen isotope record from the Norwegian-Greenland Sea that is radiocarbon-dated directly by accelerator mass spectrometry. A significant light-oxygen-isotope event is recorded at approximately 15,000 years BP which suggests that the marine-based Barents Shelf ice sheet disintegrated rapidly at this time. Recent studies^{6,7} have estimated that the decay of this ice sheet could have contributed as much as 15 metres to eustatic sea-level rise. The decay of the Barents Shelf ice sheet is the earliest major deglacial event yet dated, and may have triggered subsequent deglacial events through eustatic sea-level effects.

Boxcore PS21295-4, collected from the Fram Strait (Fig. 1) by the West German research vessel *Polarstern* in the summer of 1984, has a moderate sedimentation rate of 2.5 cm per 1,000 yr as defined by 22 accelerator mass spectrometry (AMS) ¹⁴C analyses (Table 1 and Fig. 2). With the recent development of AMS it has become possible to date directly evidence of the last deglaciation in marine sediments^{8–10}. Our core is a valuable northern endmember to those directly dated records published previously from 38° N off the coast of Portugal⁹ and 55° N off Ireland⁸. The moderate sedimentation rate, the very young core-top age (<1,000 years) and the thin mixed layer (2–3 cm) indicate that resolution of palaeoceanographic events in our isotopic record is not likely to be degraded seriously by the effects of bioturbation.

It is generally believed that the glacial/interglacial ice-volume effect changed oceanic $\delta^{18}\text{O}$ by $\sim 1.3\%$ (refs 11, 12), but the glacial-interglacial amplitude of $\delta^{18}\text{O}$ recorded in planktonic foraminifera from most regions of the world's ocean averages ~ 1.6 – 1.7% . In the North Atlantic, where isotopic records have been AMS-radiocarbon-dated previously, this amplitude approaches 2.5–3.0% (refs 8, 9). Enrichment of ^{18}O in excess of the 1.3% glacial-interglacial ice-volume signal is widely believed to be the result of the lower temperature of the sea water in which the foraminifera grew during glacial intervals. A 1°C change in temperature results in a 0.23% change in the $\delta^{18}\text{O}$ recorded in calcite precipitated in the seawater¹³, suggesting that some surface waters of the ocean were several degrees cooler during the last glaciation.

Fram Strait would be expected to preserve an isotope signal approximately equal in amplitude to the glacial-interglacial global-ice-volume signal with a minimal temperature overprint. Today this region is characterized by late-summer sea surface temperatures of $\sim 2.0^\circ\text{C}$ (ref. 14). Taking into account the freezing point of sea water and the mechanisms of sea-ice formation, this region could have experienced no more than a 3.0°C temperature depression between the last glacial maximum and today.

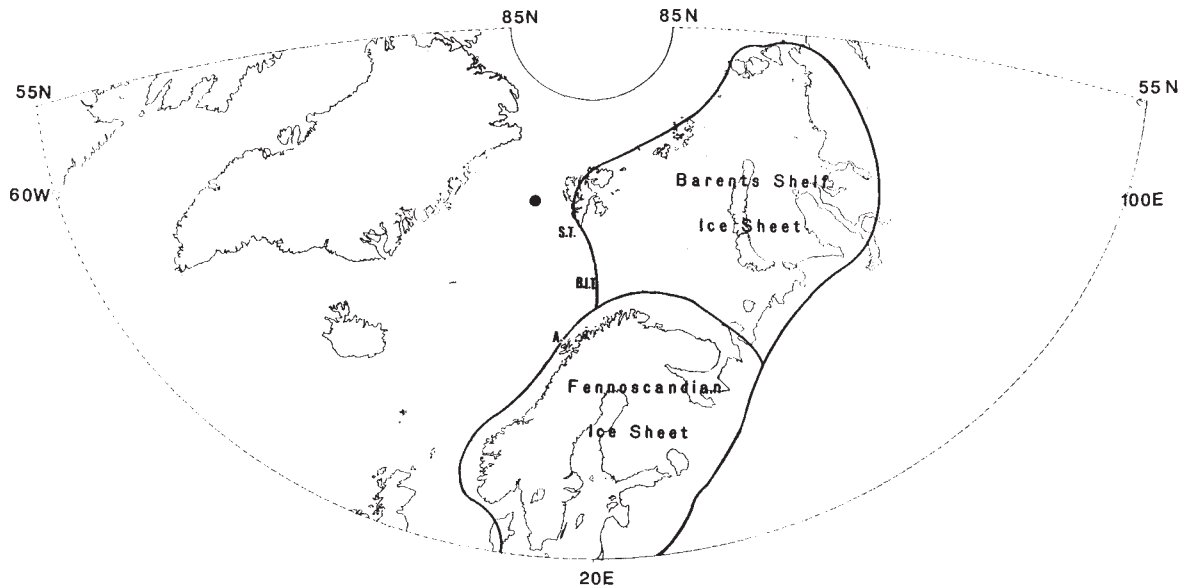


Fig. 1 Equal area projection of the Norwegian-Greenland Sea and Barents Sea. Location of core PS21295-4 (77°59.52' N, 2°25.23' E, 3112 m) is shown as a filled circle. Location of Barents Shelf and Fennoscandian ice sheets, Bear Island (BIT) and Storfjord (ST) troughs, and Andoya Island (A) are marked.

The planktonic foraminifera *Neoglobobulimina pachyderma* (left coiling) was used for both AMS radiocarbon and oxygen isotope analyses (Table 1). The isotope record is plotted in Fig. 3, based upon the results in Fig. 2 with 440 years subtracted to correct for the age difference between the marine and terrestrial radiocarbon reservoirs¹⁵. The glacial-interglacial amplitude of this record is 1.25‰, which is consistent with that expected if it were solely a record of global ice volume with no temperature overprint. We have also included in Fig. 3 the most recent estimate of the temporal variations in the global-ice-volume signal¹¹. The position of this curve (dotted line) is constrained by maximum glacial $\delta^{18}\text{O}$ values (4.7–4.9‰) for the Norwegian-

Greenland Sea, which have been obtained in this study and elsewhere^{16–18}. The midpoint of this curve is centred at 10,000 yr BP (reservoir corrected)¹¹. The isotope transition that is observed near 10,000 yr BP (termination 1b)³⁶ agrees both temporally and in amplitude with the recent estimate of global-ice-volume change. But we also record a rapid 1.1‰ $\delta^{18}\text{O}$ depletion between 36 cm (15,230 yr BP) and 34 cm (14,480 yr BP) with the midpoint of this change at ~15,000 yr BP. It is difficult to credit, and there is no evidence to suggest, the 5 °C increase in temperature at 78° N at 15,000 yr BP that would be required if this isotope anomaly was due to temperature effects alone; rather, one would expect the transfer of melt-water from the Barents Shelf and northern Fennoscandian ice sheets to the surface mixed layer of the Norwegian-Greenland Sea to be recorded at the location of core PS21295-4.

The early history of deglaciation in Scandinavia is complicated by a lack of direct measurements of, and by little agreement about, the maximum extent of the Fennoscandian ice sheet. The terrestrial record preserves no indication of the maximum extent of ice, which is believed to have extended to the edge of the continental shelf during the last glacial maximum¹⁹. There are further complications regarding the areal and volumetric extent, and even the existence, of the Barents Shelf ice sheet. Some researchers have invoked the existence of a massive ice sheet during the last glacial maximum²⁰, whereas others have argued that a small ice sheet was restricted primarily to the centre of the Svalbard archipelago^{21,22}. Other geological evidence points to the existence of an intermediate-sized ice sheet^{23–25}.

Recently, Peltier⁶ has argued that an ice mass comparable in size with the Fennoscandian ice sheet must have been located in the Barents Sea during the last glacial maximum in order to explain the observed non-tidal acceleration of rotation of the Earth and the direction of polar wander. Using evidence from the uplift rate in Svalbard and the northern Barents Sea, and the lack of a detectable marine limit on Bjornoya, Lehman²⁶ suggested either the presence of very thin ice in the western Barents Sea or an early disintegration of a more substantial Barents Shelf ice sheet. In addition, forward modelling of the record of both near- and far-field relative sea-level curves suggests that if a Barents Shelf ice sheet existed it must have disintegrated considerably earlier than 10,000 yr BP⁷.

Besides the recent modelling studies, there is also geological evidence for early deglaciation of the continental-shelf portion of the northern Fennoscandian ice sheet²⁷ at Andoya, 69.5° N

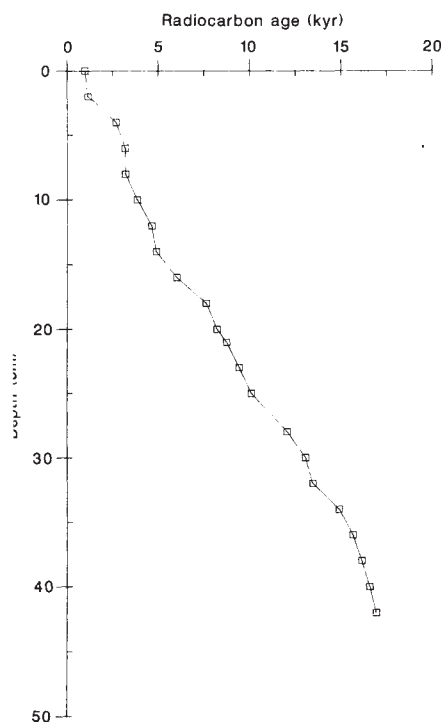


Fig. 2 Age-depth plot for core PS21295-4. Ages are not reservoir corrected. Note young core-top age (960 yr) and thin mixed layer (2–3 cm). Also, note lack of age reversals and near uniform rate of sedimentation.

Table 1 Results of oxygen isotope and radiocarbon analyses

Depth (cm)	$\delta^{18}\text{O}$ (PDB)	Accession number	Age, yr BP*	Age, yr BP (reservoir corrected)
0-1	3.46	AA-2058	960 $\pm$ 60	520
2-3	3.55	AA-2296	1,150 $\pm$ 450	710
4-5	3.69	AA-2298	2,690 $\pm$ 70	2,250
6-7	3.57	AA-2300	3,180 $\pm$ 80	2,740
8-9	3.61	AA-2302	3,190 $\pm$ 70	2,750
10-11	3.60	AA-2307	3,850 $\pm$ 60	3,410
12-13	3.61	AA-2309	4,650 $\pm$ 80	4,210
14-15	3.57	AA-2311	4,900 $\pm$ 70	4,460
16-17	3.67	AA-2313	6,020 $\pm$ 60	5,580
18-19	3.58	AA-2334	7,640 $\pm$ 80	7,200
20-21	3.68	AA-2060	8,220 $\pm$ 90	7,780
21-22	3.85	AA-2336	8,750 $\pm$ 120	8,310
23-24	3.93	AA-2338	9,440 $\pm$ 90	9,000
25-26	4.17	AA-2339	10,100 $\pm$ 100	9,660
28-29	4.35	AA-2353	12,070 $\pm$ 160	11,630
30-31	4.27	AA-2354	13,070 $\pm$ 110	12,630
32-33	4.02	AA-2355	13,480 $\pm$ 130	13,040
34-35	3.58	AA-2356	14,920 $\pm$ 120	14,480
36-37	4.65	AA-2357	15,670 $\pm$ 170	15,230
38-39	4.52	AA-2358	16,140 $\pm$ 150	15,700
39-41	4.71	AA-2359	16,580 $\pm$ 140	16,140
42-43	4.64	AA-2360	16,940 $\pm$ 170	16,500

All analyses are of *Neogloboquadrina pachyderma* (left-coiling) (150–300 micron size fraction).

* Radiocarbon ages were determined from AMS measurements made at the University of Arizona-National Accelerator Facility in March 1987. Ages were determined after C-13/C-12 normalization to -25% and using the Libby half life of 5,570 yr. Details of sample preparation and analysis are presented elsewhere^{34,35}.

off north-west Norway (Fig. 1). Although our data cannot be used to identify the exact source of the deglacial meltwater event recorded in core PS21295-4, it is reasonable to assume that the continental-shelf portion of the northern margin of the Fennoscandian ice sheet would behave synchronously with the adjacent marine-based Barents Shelf ice sheet. If this is true then our data can be used to constrain further the timing of the initiation of deglaciation of the marine-based portion of the northern Fennoscandian ice sheet.

Using a combination of the data presented here with AMS-dated isotope records from the North Atlantic^{8,9,28}, and the recent model results of Peltier^{6,29} and Nakada and Lambeck⁷, we develop a scenario for the timing of events during early deglaciation (15,000–11,000 yr BP). We assume that during the last glacial maximum, global ice volume reached an equivalent of 130 metres of eustatic sea-level fall¹². The first major event during deglaciation was the rapid destruction of the Barents Shelf ice sheet at $\sim 15,000$ yr BP. Using a simple bioturbation model³⁰ and the 2–3-cm mixed-layer thickness obtained from our AMS ^{14}C analyses, we estimate that the meltwater event, and, by inference, the decay of the Barents Shelf ice sheet, occurred over less than 500 years.

The rapidity of this decay was probably the result of the marine-based nature of the Barents Shelf ice sheet, which makes it very susceptible to decoupling of the glacier bed (caused by either a eustatically or an isostatically induced rise in sea level). Thus we view the Barents Shelf ice sheet as a self-oscillating system, for which its own increase in mass is largely responsible for its own destruction by isostatic depression of the glacier bed, and for which no external forcing is required to initiate deglaciation.

Eustatic sea level ceased falling by the last glacial maximum ($\sim 18,000$ yr BP). We suggest, however, that the Barents Shelf ice sheet continued to isostatically depress the Barents Sea floor, which is presently at an average water depth of 200 m. By 15,000 yr BP isostatic depression reached a critical level, initiating the rapid destruction of this ice sheet by marine downdraw¹

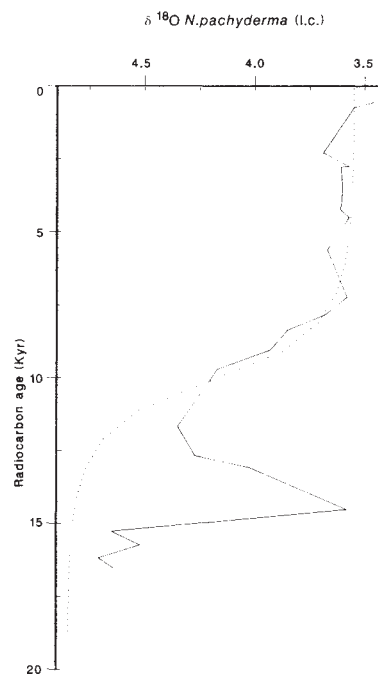


Fig. 3 Oxygen isotopic record from core PS21295-4. Approximately 40 specimens were used for each stable isotope analysis. For AMS radiocarbon analyses approximately 1,000 specimens (8–12 mg) of *N. pachyderma* (left-coiling) were isolated from the core, converted to graphite and analysed at the University of Arizona accelerator facility according to procedures outlined elsewhere^{34,35}. Although it is possible that the apparent radiocarbon age of the last glacial maximum oceanic reservoir was different from the 440-yr reservoir correction for the modern Norwegian-Greenland Sea¹⁵, there is no direct evidence at present to support or refute variations of apparent age back in time. Thus we have assumed that the 440-yr reservoir correction applies over the entire interval of time spanned by core PS21295-4. The dotted line is the global ice volume estimated by Ruddiman¹¹. Note significant $\delta^{18}\text{O}$ minimum near 15,000 yr BP.

through the Bear Island and Storfjord troughs (Fig. 1) and forming large depositional fans at the mouths of these troughs³¹. At this time the Norwegian-Greenland Sea was still largely isolated from the North Atlantic, as meridional transport was not initiated until $\sim 13,000$ yr BP³². The timing of the discharge of isotopically light ice and meltwater into the Norwegian-Greenland Sea surface mixed layer is preserved as the isotopically light event that we have dated in core PS21295-4 at $\sim 15,000$ yr BP. Although there are too many uncertainties to allow us to estimate the size of this ice sheet from the amplitude of the isotopic light event recorded in Fram Strait, its size has been implicitly determined by Nakada and Lambeck⁷ and Peltier⁶ to be equivalent to ~ 15 m of eustatic sea-level rise. We speculate that the susceptibility of the marine-based Barents Shelf ice sheet to rapid destruction by marine down-draw acts as a triggering mechanism for subsequent decay of portions of the Laurentide ice sheet around Cabot Strait¹¹ through eustatic sea-level effects.

Finally, the recent isostatic-adjustment modelling studies of Peltier²⁹ and Nakada and Lambeck⁷ also require the timing of the decay of the Antarctic ice sheet to be delayed from $\sim 18,000$ yr BP³³ to $\sim 11,000$ yr BP. In combination with our data for early deglaciation in the Northern Hemisphere, these results suggest that deglacial events in the Northern Hemisphere may have triggered the rapid decay of the Antarctic ice sheet, rather than early decay of the Antarctic ice sheet triggering a deglaciation of the Northern Hemisphere ice sheets through eustatic sea-level effects.

We thank J. Thiede for providing a subcore of PS21295-4, S. Lehman for stimulating discussions, A. J. T. Jull, D. Donahue and T. Linick for help with the AMS analyses, E. Franks for stable isotope analyses and K. Elder and L. Toolin for assistance in sample preparation. This work was supported by the NSF and ONR. The University of Arizona-National Accelerator Facility is supported by the NSF.

Received 13 June; accepted 5 September 1988.

1. *The Last Great Ice Sheets* (eds Denton, G. H. & Hughes, T. J.) (Wiley Interscience, New York, 1981).
2. Ruddiman, W. F. & Duplessy, J. C. *Quat. Res.* **23**, 1-17 (1985).
3. *Quaternary Glaciations in the Northern Hemisphere* (eds Sibrava, V., Bowen, D. Q. & Richmond, G. M.) (Pergamon, Oxford, 1986).
4. *North America and Adjacent Oceans During the Last Deglaciation* (eds Ruddiman, W. F. & Wright, H. E. Jr) (Geol. Soc. Am., Boulder, 1987).
5. *The Laurentide Ice Sheet: Geogr. phys. Quat.* **41** (eds Fulton, R. J. & Andrews, J. T.) (1987).
6. Peltier, W. R. *Science* **240**, 895-901 (1988).
7. Nakada, M. & Lambeck, K. *Nature* **333**, 36-40 (1988).
8. Duplessy, J. C. *et al. Nature* **320**, 350-352 (1986).
9. Bard, E., Arnold, M., Duprat, J., Moyes, J. & Duplessy, J. C. *Nature* **328**, 791-794 (1987).
10. Broecker, W. S. *et al. Paleoceanography* **3**, 1-19 (1988).
11. Ruddiman, W. F. in *North America and Adjacent Oceans During the Last Deglaciation* (eds Ruddiman, W. F. & Wright, H. E., Jr) 463-478 (Geol. Soc. Am. Boulder, 1987).
12. Chappell, J. & Shackleton, N. J. *Nature* **324**, 137-140 (1986).
13. Epstein, S., Buchsbaum, R., Lowenstam, H. A. & Urey, H. C. *Bull. geol. Soc. Am.* **64**, 1315-1328 (1953).
14. Aagaard, K., Goldvik, A. & Hillman, S. R. *J. geophys. Res.* **92**, 3778-3784 (1987).
15. Mangerud, J. & Gulliksen, S. *Quat. Res.* **5**, 263-273 (1975).
16. Kellogg, T. B., Duplessy, J. C. & Shackleton, N. J. *Boreas* **7**, 61-73 (1978).
17. Zahn, R., Markussen, B. & Thiede, J. *Nature* **314**, 433-435 (1985).
18. Sejrup, H. P., Jansen, E., Erlenkeuser, H. & Høltedahl, H. *Quat. Res.* **21**, 74-84 (1984).
19. Andersen, B. G. in *The Last Great Ice Sheets* (eds Denton, G. H. & Hughes, T. J.) 1-65 (Wiley Interscience, New York, 1981).
20. Grosswald, M. G. *Quat. Res.* **13**, 1-32 (1980).
21. Boulton, G. S. *Boreas* **8**, 31-57 (1979).
22. Boulton, G. S. *Boreas* **8**, 373-395 (1979).
23. Elverhoi, A. & Solheim, A. *Polar Res.* **1**, 23-42 (1983).
24. Vorren, T. O. & Kristofferson, Y. *Boreas* **15**, 51-59 (1986).
25. Forman, S. L., Mann, D. H. & Miller, G. H. *Quat. Res.* **27**, 41-50 (1987).
26. Lehman, S. J. thesis, Univ. of Colorado, Boulder (1987).
27. Vorren, T. O., Vorren, K.-D., Alm, T., Gulliksen, S. & Lovlie, R. *Boreas* **17**, 41-77 (1988).
28. Keigwin, L. D. & Jones, G. A. *Deep Sea Res.* (submitted).
29. Peltier, W. R. *Quat. Res.* **29**, 93-112 (1988).
30. Berger, W. H. & Heath, G. R. *J. mar. Res.* **26**, 134-142 (1968).
31. Solheim, A. & Kristofferson, Y. Norsk Polarinstitutt, Skriftr NR 179B (1984).
32. Jansen, E. & Björklund, K. R. *Boreas* **14**, 243-257 (1985).
33. Clark, J. A. & Lingle, C. S. *Quat. Res.* **11**, 279-298 (1979).
34. Slota, P. J. Jr, Jull, A. J. T., Linick, T. W. & Toolin, L. J. *Radiocarbon* **29**, 303-306 (1987).
35. Jull, A. J. T., Donahue, D., Hatheway, A. L., Linick, T. W. & Toolin, L. J. *Radiocarbon* **28**, 191-197 (1986).
36. Duplessy, J. C., Delibrias, G., Turon, J. L., Pujol, C. & Duprat, J. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **35**, 121-144 (1981).

Wave transport of deep mantle material

John A. Whitehead & Karl R. Helfrich

Department of Physical Oceanography,
Woods Hole Oceanographic Institution,
Woods Hole, Massachusetts 02543, USA

Fluid with a certain density and viscosity can rise by buoyant Poiseuille flow through a conduit¹ within a second fluid of greater density and viscosity. Such conduits exhibit a rich behaviour characteristic of nonlinear systems, an aspect of which is the formation of solitary waves^{2,3}. Here we present theoretical and experimental studies of these systems. Both approaches reveal that solitary waves trap material in a cell with closed streamlines and that the central streamline velocity is faster than the wave speed. Hence, parcels of deep material are transported directly upward over large distances. This is in contrast to the usual situation in which wave propagation through a medium causes only small displacement of fluid particles. Material in these parcels will be far less contaminated by diffusion from the surroundings than would be material in ordinary pipe flow. In addition, solitary waves are more efficient than buoyant spheres at conveying material upward. We suggest that such waves might exist in the Earth's mantle, conveying uncontaminated deep mantle material to the surface of the Earth.

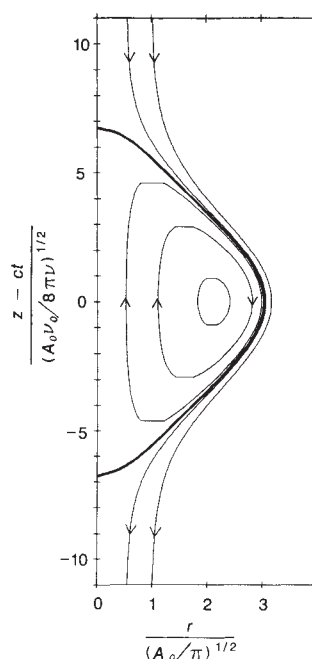


Fig. 1 Streamlines of a solitary wave calculated in the frame moving with the solitary wave speed c for the case $\varepsilon = 9$. Radius r is normalized by the radius of the conduit and the conduit axial dimension z is normalized by a natural length scale of the problem³. Note the large region of closed streamlines.

There has been considerable recent interest in buoyant conduits, for several reasons. First, geological and geophysical counterparts may exist in the Earth: applications have been made to magma chambers⁴, mantle convection^{5,6} and hotspots^{7,8}. Seismic evidence shows that the core-mantle boundary and the associated D'' layer (which can be interpreted as a thermal boundary layer) is rough. Numerical simulations⁹ of mantle convection in a fluid with a strongly temperature-dependent viscosity show conduits in which unstable solitary waves are generated. Second, the system is a simple analogue of compaction-driven flow in a porous viscous matrix², which bears upon such diverse issues as the formation of melt under mid-ocean spreading centres, the formation of melt under hotspot islands such as Hawaii or Iceland, and the nature of mud volcanoes and diapirs. Scott¹⁰ noted that solitary waves in the compaction problem do not convey material over large distances because the waveform ascends faster than the background flow. Thus our new findings distinguish the conduit problem from its compaction analogue.

The clearest demonstration of closed streamlines are theoretical. In the limit of small amplitude, characterized by $\varepsilon \equiv A_m/A_0 - 1 \ll 1$, the wave speed is $c = (g' A_0/4\pi\nu)[1 + \varepsilon/3]$, whereas the maximum (centre-line) vertical fluid velocity in the conduit is $u_m = (g' A_0/4\pi\nu)[1 + \varepsilon]$, so $u_m > c$. Here A_m is the maximum area or amplitude of the wave, A_0 is the area of the undisturbed conduit, $g' = g\Delta\rho/\rho$ is gravity multiplied by normalized density difference between exterior and conduit fluids and ν is the viscosity of the conduit fluid. In a frame co-moving with the wave, fluid moves forward at the exact centre of the wave and backward in the undisturbed pipe (both upstream and downstream), and stagnation points are found both in front of and behind the centre of the wave. Fluid is trapped in closed streamlines between these two points and conveyed with the wave. The above argument is valid for small ε , but this result is found at arbitrary wave amplitude. Figure 1 shows theoretical streamlines for $\varepsilon = 9$ (K.R.H. and J.A.W., preprint).

We have performed simple laboratory experiments which clearly demonstrate fluid trapping. The apparatus is the same as described previously⁹ and essentially consists of a glass vessel

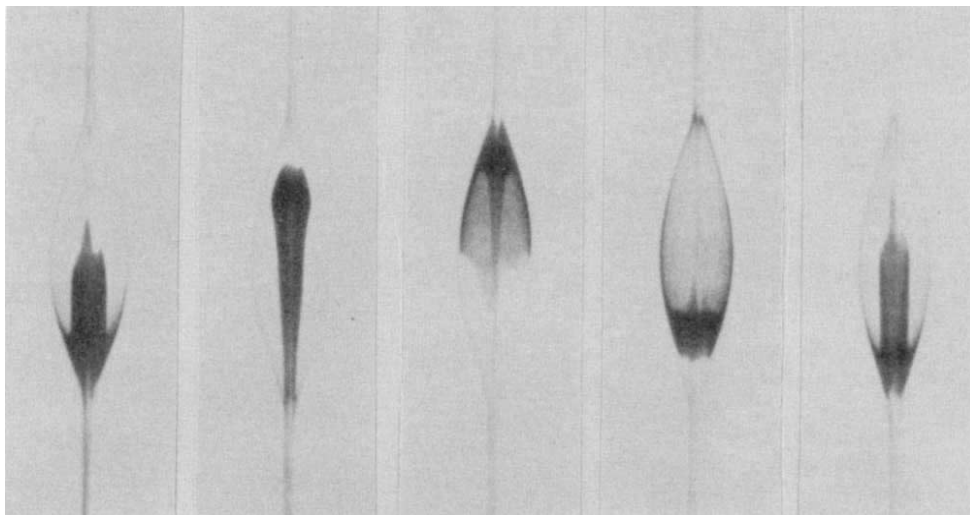


Fig. 2 Sequential photographs (with equal time intervals) of a solitary wave with a small amount of dye. This reveals the circulation in closed streamlines. The conduit and most of the wave contains clear fluid which is scarcely distinguishable.

containing corn syrup with a tubular source of syrup-water mixture on the bottom. Figure 2 contains a time series of photographs that follow a single solitary wave. The conduit and much of the wave are made up of clear fluid which is difficult to distinguish from the external fluid. Some dyed fluid was injected into the wave, and circulates within the wave by ascending along the centre-line and descending close to the periphery. On reaching the bottom of the wave, it repeats the circuit. These photographs indicate that the fluid is trapped within the waves and that the circulation pattern is qualitatively similar to the theory (Fig. 1). Clear observations of trapped fluid have also been made in waves of very small amplitude and in rank ordered packets of waves of extremely large amplitude⁹.

If conduits extend upward from some great depth in the Earth's mantle, they might contain such waves in response to changes in volume flow from transient boundary conditions. They will then convey isolated packets of deep mantle material upward in their centres and deposit the material in the upper mantle, possibly under hotspots. The following calculations indicate that the dilution of deep mantle material from sideways diffusion in waves would be quite small relative to the corresponding dilution in ordinary pipe flow.

We consider first the dilution in pipe flow. Taylor¹¹ showed that matter dispersed in a solvent in pipe flow possesses a virtual coefficient of diffusivity $D_v = a^2 u_0^2 / 192D$ in the direction of pipe flow, where a is pipe radius, u_0 is the maximum velocity at the axis and D is the molecular diffusivity of the material. This relation is valid when $4L/a \gg u_0 a / D \gg 6.9$ (ref. 12), with L the conduit length. If we use for typical values¹³ a mantle conduit radius $a = 5$ km, conduit length $L = 3,000$ km (depth of mantle), molecular diffusivity $D = 10^{-8} \text{ m}^2 \text{ s}^{-1}$ and a velocity $u_0 = 10^{-9} \text{ m s}^{-1}$ (equivalent to a rate of 3 cm yr^{-1} , which is the spreading rate of moderately fast tectonic plates and gives a rise of $3,000$ km in 100 Myr) then the inequalities are satisfied. Consider a pulse of new material injected in a conduit undergoing waveless pipe flow. The concentration anomaly of the new material can be represented as a gaussian function which diffuses vertically with width $\sim \sqrt{D_v t}$. The pulse amplitude decreases as the inverse of this width. If the initial width is equal to the conduit radius a and rises through the mantle from the core, it will have been diluted with surrounding conduit material by an amount

$$\frac{a}{\sqrt{D_v L / u_0}} = \sqrt{\frac{192D}{Lu_0}} = 2.5 \times 10^{-2}$$

which might make it undetectable in upper mantle material. Using a smaller value of D leads to even more dilution, although one of the inequalities begins to be violated. We infer, therefore,

that in pipe flow, materials would get strongly dispersed as they rose through a conduit in the mantle.

One reason for the extensive mixing of material in pipe flow is that the surface area between the material and surrounding fluid increases with time. In contrast, material confined within a wave with closed streamlines will retain a constant surface area. The law governing the evolution of concentration C in a well mixed volume V with surface area S as it rises through a conduit of thickness d is

$$V \frac{dC}{dt} = -\frac{SD}{d} C$$

with solution concentration $C = C_0 e^{-SDt/Vd}$. Using the values $t = L/u_0 = 100$ Myr and thickness of both the wave and the boundary layer $V/S = d = 5$ km, along with the previous value of D , the concentration will decrease by $\sim e^{-1.2} = 0.3$. With such a dilution, a substance originating from the core-mantle interface or at some other deep mantle source might be chemically or isotopically unaltered by the time it reaches the surface of the Earth. We note that hotspot basalt has different trace-element and isotope signatures from mid-ocean ridge basalt¹⁴, indicating a lack of dilution of the hotspot material between its source and the Earth's surface.

Waves in conduits are also a preferable mode of vertical mantle transport because they transport material more efficiently than spherical cavities. We can estimate the size of a wave required for it to rise through the mantle in a reasonable geological timescale. Olson and Christensen³ show that the speed of large solitary waves is (in dimensional form)

$$c = \left(\frac{g'a^2}{4\nu} \right) \left(\ln \left(\frac{A_m}{A_0} \right) - \frac{1}{2} \right)$$

To obtain for c the geophysically relevant value of 10^{-9} m s^{-1} , with the parameter $g' = 0.01 \text{ m s}^{-2}$, $\nu = 10^{14} \text{ m}^2 \text{ s}^{-1}$ (assuming conduit viscosity is 10^{-4} of the commonly accepted mantle estimate $\nu_0 = 10^{18} \text{ m}^2 \text{ s}^{-1}$ (ref. 13)) and $a = 5.0$ km, we get $A_m/A_0 = 8.17$, which implies that the wave radius $a_m = 14.3$ km. The volume of material in a wave is then

$$\Delta V = \frac{\pi a^3 A_m}{A_0} \sqrt{\left(\frac{\pi \nu_0}{2\nu} \right) \left(\ln \left(\frac{A_m}{A_0} \right) - \frac{1}{2} \right)} = 3.9 \times 10^5 \text{ km}^3$$

The volume flux of the base conduit is

$$Q_c = \frac{A_0^2 g'}{8\pi\nu} = 7.7 \times 10^5 \text{ m}^3 \text{ yr}^{-1}$$

Dividing ΔV by Q_c , we see that one solitary wave of this type every 500 million years would double the net volume flux from

the conduit. Thus, wave transport is a very efficient means of increasing net conduit volume flux. In contrast, a sphere must be significantly larger to rise through the mantle in geological time. The sphere velocity v_s is¹

$$v_s = \frac{g'r_s^2}{3\nu_0}$$

To give the same velocity as before, we require a radius $r_s = 550$ km, and therefore a much greater volume of 7.0×10^8 km³. Spheres are inefficient for other reasons as well. Griffiths¹⁵ has shown that spheres smaller than 200 km are too slow ever to reach the surface.

Thus we envisage a process of mantle flow in which solitary waves trap parcels of material and easily transport them vertically with little dilution. Note that mantle (especially lower mantle) characteristics are poorly constrained, with uncertainties much greater than factors of ten, so the estimates of diffusion velocities are only suggestive. We have also ignored thermal and multicompositional effects, although thermal effects are modelled well by laboratory compositional experiments. But the description presented here avoids problems associated with very wide mantle plumes of uniform viscosity or with very large spheres in the mantle.

Received 14 July; accepted 12 September 1988.

1. Whitehead, J. A. & Luther, D. S. *J. geophys. Res.* **80**, 705-717 (1975).
2. Scott, D. S., Stevenson, D. L. & Whitehead, J. A. *Nature* **319**, 759-761 (1986).
3. Olson, P. & Christensen, U. *J. geophys. Res.* **91**, 6367-6374 (1986).
4. Huppert, H. E., Sparks, R. S. J., Whitehead, J. A. & Hallworth, M. A. *J. geophys. Res.* **91**, 6113-6122 (1986).
5. Stacey, F. D. & Loper, D. E. *Phys. Earth planet. Inter.* **33**, 45-55 (1983).
6. Olson, P., Schubert, G. & Anderson, C. *Nature* **327**, 409-412 (1987).
7. Skilbeck, J. N. & Whitehead, J. A. *Nature* **272**, 499-501 (1978).
8. Whitehead, J. A. *Geophys. J. R. astr. Soc.* **70**, 415-433 (1982).
9. Whitehead, J. A. *Am. J. Phys.* **55**, 998-1003 (1987).
10. Scott, D. S. thesis, California Institute of Technology, California (1987).
11. Taylor, G. I. *Proc. R. Soc. A* **219**, 186-203 (1953).
12. Taylor, G. I. *Proc. R. Soc. A* **225**, 473-477 (1954).
13. Turcotte, D. L. & Schubert, G. *Geodynamics* (Wiley, New York, 1982).
14. Hofmann, A. W. *Terra Cognita* **4**, 157-165 (1984).
15. Griffiths, R. W. *Earth planet. Sci. Lett.* **78**, 435-446 (1986).

Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*

Jeffrey L. Feder, Charles A. Chilcote* & Guy L. Bush

Department of Zoology, Michigan State University, E. Lansing, Michigan 48824, USA

* Department of Natural Resources, University of Michigan, Ann Arbor, Michigan 48109, USA

It has been proposed that true fruit flies in the *Rhagoletis pomonella* species group speciate sympatrically (that is, in the absence of geographic isolation) as a consequence of shifts to previously unexploited host plants^{1,2}. Because *Rhagoletis* larvae are host-specific fruit parasites and mate selection in these flies is directly coupled to host-plant recognition³⁻⁵, variation for larval survivorship and host preference traits can act as genetically based barriers to gene flow. This reduction in gene flow results in the sympatric divergence of fly populations adapted to alternative hosts. The shift of *R. pomonella* from its native host hawthorn (*Crataegus* spp.) to domestic apples (*Malus pumila*) in eastern North America in the past 200 years⁶ provides an opportunity to determine whether host specialization is sufficient to differentiate populations without earlier periods of geographic isolation. We report finding genetic differentiation between co-occurring hawthorn and apple populations of *R. pomonella* at a field site near Grant, Michigan. The result confirms that hawthorn and apple flies represent partially reproductively isolated 'host races' and is consistent with a sympatric mode of divergence for these flies.

Six allozyme loci, aconitase-2 (*Acon-2*), malic enzyme (*Me*), mannose phosphate isomerase (*Mpi*), aspartate amino-transferase (*Aat-2*), NADH-diaphorase-2 (*Dia-2*) and beta-hydroxy acid dehydrogenase (*Had*), showed significant allele frequency differences between flies collected from sympatric hawthorn and apple trees at the Grant, Michigan site (Tables 1 and 2; see Fig. 1 and accompanying legend for a description of the field site, collecting strategy and electrophoretic methods used in the study). Allele frequencies for *Acon-2* and *Me* were significantly different between hawthorn and apple flies across the three years sampled and did not vary significantly among either apple or hawthorn trees within any year (Tables 1 and 2). *Aat-2*, *Dia-2* and *Mpi* also showed host-related differences. But gene frequencies varied among flies from different apple trees for *Aat-2* in 1985 and for *Dia-2* in 1984 and 1985, as well as among hawthorn trees for *Mpi* in 1985 (Tables 1 and 2). Allele frequencies for *Had* differed significantly between hawthorn and apple flies in 1985 but not in 1984 or 1986. *Had* also varied significantly among apple trees in 1984. None of the other seven polymorphic loci resolved in this study showed significant inter- or intra-host frequency differences.

Allele frequencies were reasonably constant between 1984 and 1985. Of the 13 polymorphic loci, only apple tree number 1 for *Me* ($G = 9.67$, $P \leq 0.01$, 1 d.f.) and hawthorn tree number 1 for *Dia-2* ($G = 4.66$, $P \leq 0.05$, 1 d.f.) had significant frequency differences between 1984 and 1985 (data for *Acon-2*, *Me*, *Mpi*, *Aat-2*, *Dia-2* and *Had* are presented in Table 2). But eight out of a total of 20 tests for *Me*, *Acon-2*, *Mpi*, *Aat-2* and *Dia-2* showed significant differences between 1985 and 1986. Allele frequencies for both apple and hawthorn flies displayed a general trend to become more 'apple-like' in 1986 compared with previous years (Table 1). The shift in 1986 allele frequencies could be explained by selection either in the field or during laboratory rearing, as flies collected as larvae in 1985 and adults in 1986 represent different life-history stages of the same generation. But frequency shifts between 1985 and 1986 were in the same direction and of similar magnitude for both hawthorn and apple flies (Table 1). Therefore, if selection did occur, its effects were uniform across host plants.

Allele frequencies for *Acon-2*, *Me*, *Mpi*, *Aat-2*, *Dia-2* and *Had* may not be evolving independently. Hitch-hiking effects for linked loci and epistatic selection can coordinately change allele frequencies at a number of loci. To examine these possibilities, correlation coefficients were calculated between pairs of non-allelic genes based on Burrows disequilibrium values^{7,8}. Significant disequilibrium was found between *Me/Acon-2*, *Me/Mpi*, *Acon-2/Mpi* and *Aat-2/Dia-2* (Table 3). These results agree with the known genetic map of *R. pomonella* as *Aat-2* and *Dia-2* are 3.2 centimorgans apart on linkage group I whereas *Acon-2*, *Me* and *Mpi* are very tightly linked on linkage group II (Feder *et al.*, manuscript in preparation). Loci showing genetic differentiation between hawthorn and apple flies therefore map to three different regions of the genome, and hitch-hiking effects are probable among non-allelic genes within at least two of these regions.

Interactions between unlinked loci were not readily apparent from the correlation coefficients. Twenty-one significant tests of gametic disequilibrium were observed in 314 pairwise comparisons between *Mpi*, *Acon-2* or *Me* and either *Aat-2* or *Dia-2* (data not shown). *Had*, which has been mapped to linkage group III (ref. 9), was in gametic equilibrium with either *Acon-2*, *Me*, *Mpi*, *Aat-2* or *Dia-2* in 153 of 160 tests. Although interchromosomal levels of gametic disequilibrium cannot be completely explained by random type I errors, no consistent pattern was observed in either the loci or alleles involved in the significant tests. Allele frequencies for unlinked loci therefore appear to be evolving independently in both hawthorn and apple populations.

Several factors, alone or in combination, could be responsible for the observed allele frequency differences between hawthorn

Table 1 Allele frequencies for *Me*, *Acon-2*, *Mpi*, *Aat-2*, *Dia-2* and *Had* for flies collected from hawthorn and apple trees at the Hansen site near Grant, Michigan

Year (Stage)	Host	<i>Me</i>		<i>Acon-2</i>			<i>Mpi</i>			<i>Aat-2</i>			<i>Dia-2</i>		<i>Had</i>	
		<i>n</i>	80	<i>n</i>	100	95	<i>n</i>	100	37	<i>n</i>	100	21	<i>n</i>	100	<i>n</i>	100
1984 (Larvae)	Apple-1	99	0.707	99	0.545	0.152	99	0.889	0.056	99	0.279	0.187	96	0.688	99	0.803
	Apple-2	84	0.631	86	0.512	0.221	86	0.884	0.046	86	0.326	0.157	86	0.767	86	0.907
1985 (Larvae)	Apple-1	25	0.480	25	0.480	0.240	25	0.900	0.060	25	0.300	0.120	24	0.729	25	0.740
	Apple-2	52	0.587	62	0.484	0.185	65	0.869	0.085	52	0.393	0.183	63	0.802	65	0.838
	Apple-3	51	0.618	50	0.511	0.220	50	0.880	0.060	51	0.294	0.157	50	0.650	51	0.726
1986 (Adults)	Apple-1	46	0.663	46	0.598	0.109	46	0.946	0.000	46	0.272	0.130	46	0.669	46	0.837
	Apple-2	39	0.603	39	0.584	0.167	39	0.897	0.000	39	0.269	0.141	39	0.628	39	0.859
1984 (Larvae)	Haw-1	101	0.272	101	0.297	0.490	101	0.703	0.178	101	0.465	0.104	100	0.895	101	0.876
1985 (Larvae)	Haw-1	52	0.269	64	0.250	0.500	65	0.785	0.085	51	0.372	0.088	64	0.797	65	0.885
	Haw-2	53	0.274	53	0.273	0.500	53	0.717	0.216	53	0.415	0.066	53	0.877	53	0.877
	Haw-3	51	0.225	52	0.250	0.645	52	0.683	0.183	50	0.410	0.030	52	0.865	52	0.875
1986 (Adults)	Haw-1	61	0.385	61	0.443	0.410	61	0.795	0.148	61	0.402	0.090	61	0.820	61	0.813
	Haw-2	63	0.444	63	0.365	0.365	63	0.787	0.131	63	0.460	0.048	63	0.802	63	0.841

The table is abbreviated and contains the subset of alleles with the highest levels of inter-host differentiation. Alleles were numbered according to their relative mobility to the common allele (100) at the locus. *n*, sample size. Allele *Acon-2*⁹⁵ in this study is the same as *Acon-2*⁹¹ in McPherson *et al.*¹⁶.

Table 2 G-contingency tests for allele frequency differences for *Me*, *Acon-2*, *Mpi*, *Aat-2*, *Dia-2* and *Had* at the Grant, Michigan site

Year (Stage)	Test	<i>Me</i>	<i>Acon-2</i>	<i>Mpi</i>	<i>Aat-2</i>	<i>Dia-2</i>	<i>Had</i>
1986 (Adults)	Among apple trees	0.67 (1)	1.44 (3)	1.39 (2)	5.72 (4)	0.86 (1)	0.16 (1)
	Among haw trees	0.90 (1)	6.71 (3)	0.65 (2)	2.34 (4)	0.13 (1)	0.61 (1)
	Between apple-haw	19.90 (1)*	34.60 (3)*	38.10 (2)*	14.79 (4)†	11.32 (1)*	0.44 (1)
1985 (Larvae)	Among apple trees	3.07 (2)	3.71 (6)	0.96 (4)	16.45 (8)‡	6.55 (2)‡	4.90 (2)
	Among haw trees	0.78 (2)	10.67 (6)	11.84 (4)‡	10.98 (8)	3.33 (2)	0.06 (2)
	Between apple-haw	60.70 (1)*	44.10 (3)*	—	—	—	10.97 (1)*
	Apples-haw tree 1	—	—	7.10 (2)‡	—	—	—
	Apples-haw tree 2	—	—	15.50 (2)*	—	—	—
	Apples-haw tree 3	—	—	18.50 (2)*	—	—	—
	Apple tree 1-haws	—	—	—	5.43 (4)	3.47 (1)	—
	Apple tree 2-haws	—	—	—	21.22 (4)*	1.11 (1)	—
	Apple tree 3-haws	—	—	—	16.93 (4)†	16.36 (1)*	—
1984 (Larvae)	Among apple trees	2.39 (1)	6.32 (3)	1.13 (2)	2.17 (4)	4.14 (1)‡	8.12 (1)†
	Between apple-haw	85.70 (1)*	59.30 (3)*	32.40 (2)*	31.10 (4)*	—	—
	Apple tree 1-haws	—	—	—	—	26.42 (1)*	4.01 (1)‡
	Apple tree 2-haws	—	—	—	—	9.11 (1)†	0.91 (1)
	Apple tree 1	5.01 (1)‡	6.30 (3)	6.47 (2)‡	2.93 (4)	0.17 (1)	1.87 (1)
1985/1986 (Larvae/Adults)	Apple tree 2	0.05 (1)	3.19 (3)	12.60 (2)†	17.98 (4)†	7.30 (1)†	0.16 (1)
	Haw tree 1	3.43 (1)	15.50 (3)†	5.25 (2)	0.88 (4)	0.21 (1)	3.20 (1)
	Haw tree 2	7.33 (1)†	9.39 (3)‡	3.00 (2)	6.52 (4)	2.45 (1)	0.62 (1)
	Apple tree 1	9.67 (1)†	5.60 (3)	0.04 (2)	3.75 (4)	0.62 (1)	0.92 (1)
1984/1985 (Larvae)	Apple tree 2	1.45 (1)	2.72 (3)	1.22 (2)	5.52 (4)	0.50 (1)	3.19 (1)
	Haw tree 1	0.16 (1)	4.85 (3)	5.97 (2)	4.12 (4)	4.66 (1)‡	0.05 (1)

Degrees of freedom are given in parentheses. Gene frequencies for non-significant intra-host tests were pooled across host trees and these pooled totals were used for inter-host tests (between apple-haw). When significant intra-host heterogeneity was observed for a locus, inter-host tests were conducted on a tree-by-tree basis against the pooled total for the other host (namely apples-haw tree 1).

* $P \leq 0.001$; † $P \leq 0.01$; ‡ $P \leq 0.05$.

and apple flies. Among the possibilities are: (1) post mating reproductive isolation between hawthorn and apple flies; (2) a genetic bottleneck associated with the founding of the apple race; (3) differential larval survivorship associated with the host fruit environment; (4) differential host recognition by adult flies; (5) temporal differences in the timing of adult emergence.

It is very unlikely that post-mating reproductive isolation is involved in differentiating host populations. Hawthorn and apple flies readily mate in the laboratory and produce viable F_1 progeny (R. Prokopy, personal communication). Per cent egg hatch is similar for apple $\times$ apple, hawthorn $\times$ hawthorn and hawthorn $\times$ apple test crosses¹⁰. Although data from F_2 and backcrosses are needed to completely rule out postmating isolation, the likelihood of reproductive incompatibility between these flies is remote.

A genetic bottleneck during the colonization(s) of apple by *R. pomonella* may have initially caused gene frequency differen-

ces between the host races through drift. But for neutral genetic differences to persist requires restricted gene flow between sympatric apple and hawthorn populations. Variation for intrinsic biological factors (that is, differential larval survivorship, host recognition and/or adult emergence) are therefore necessary to account for the continued maintenance of host race differentiation regardless of its original cause.

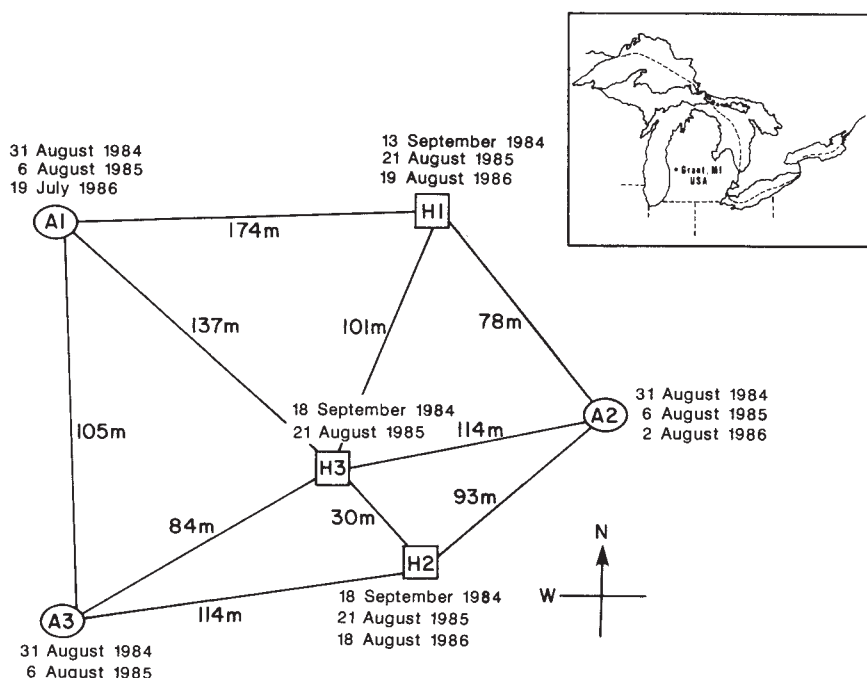
Although differential selection on larval populations infesting hawthorn and apple fruits could account for genetic divergence, larval selection alone cannot explain the significant frequency differences observed between adults captured directly from hawthorn and apple fruits in 1986 (Tables 1 and 2). Random dispersal of flies between host plants would homogenize adult gene frequencies even if larval selection was intense¹¹. Furthermore, adult hawthorn and apple flies show the same pattern and magnitude of host related differentiation as larvae. Gene flow must therefore be restricted between hawthorn and apple popu-

Table 3 Correlation coefficients (r^{AB}) calculated between pairs of non-allelic genes based on disequilibrium values^{7,8}

Year (Stage)	Host	$Me^{100}/Acon^{95}$	Me^{100}/Mpi^{100}	$Mpi^{100}/Acon^{95}$	$Aat-2^{21}/Dia-2^{85}$
1984 (Larvae)	Apple 1	0.3031*	-0.2823‡	-0.3137*	0.5045*
	Apple 2	0.4242*	-0.1276	-0.1755	0.4752*
	Haw 1	0.2770†	-0.1980‡	-0.3039†	0.6692*
1985 (Larvae)	Apple 1	0.2327	-0.3770‡	-0.4354*	0.6991†
	Apple 2	0.3314‡	-0.1997	-0.1782	0.2935‡
	Apple 3	0.5097*	-0.1833	-0.2660	0.6337†
	Haw 1	0.4634‡	-0.2780‡	-0.2285	0.5082*
	Haw 2	0.3723†	-0.2731	-0.1681	0.5153*
	Haw 3	0.3951‡	-0.3420‡	-0.2488	0.5357*
1986 (Adults)	Apple 1	0.3311†	-0.1679	-0.3311‡	0.5513*
	Apple 2	0.3229†	-0.2724	-0.0362	0.4481*
	Haw 1	0.4551†	-0.3445†	-0.3017‡	0.4990*
	Haw 2	0.5882*	-0.1853	-0.4217*	0.5267*

Significance determined by χ^2 goodness-of-fit tests of Burrows value.* $P \leq 0.001$; † $P \leq 0.01$; ‡ $P \leq 0.05$.

Fig. 1 Scale diagram of study site located near Grant, Michigan, United States. The site is an old field which has remained fallow since at least 1922 (E. Hansen, personal communication). Host tree designations (A, apple; H, hawthorn) are given with collecting dates and inter-tree distances (in metres). In 1984 and 1985 flies were sampled as larvae from infested fruit and reared to adulthood in the laboratory. In contrast, adults were captured directly from host fruits by sweep net in 1986. Standard horizontal starch gel electrophoretic techniques were used¹⁷ and a total of 29 different allozyme systems were resolved (a list of resolved systems is available on request). Isozyme systems that migrated to nearest the cathode were designated 1, the second nearest 2, and so on. Thirteen loci were polymorphic (frequency of the common allele <0.95) and, excluding *Aat-1* which is sex linked, only 15 significant deviations from Hardy-Weinberg equilibrium were observed for these loci in 290 tests. The significant deviations displayed no regular pattern and their number does not differ from random expectation because of type I error.



lations, indicating that the host races are not mating randomly in nature.

Traits involved in host recognition and the timing of adult eclosion are the most likely candidates responsible for reducing gene flow between hawthorn and apple races. Host acceptance behaviours do differ between hawthorn and apple flies¹² and former adult experience can affect host preference in *R. pomonella*¹³. Field studies have shown that hawthorn and apple flies differ in their mean adult emergence times by about a week and a half in the field¹⁴ and by even more under laboratory conditions¹⁵. Also, the distribution of adults at the Grant site closely follows the fruiting phenology of their host plants, because early in the summer flies are abundant on ripe apples but scarce on immature hawthorns, whereas the reverse is true three to four weeks later in the season when hawthorns ripen (see Fig. 1 for adult collecting dates). Differences in diapause termination and host preference, together with adult conditioning, could all contribute to the establishment of an assortative mating system, with early emerging adults tending to reproduce on apples and later emerging flies on hawthorns.

The existence of genetic differentiation between sympatric hawthorn and apple populations of *R. pomonella* at the Grant,

Michigan site confirms the status of these flies as non-randomly mating host races. The recent formation of the apple race indicates that these flies are diverging sympatrically in the absence of geographic isolation. More research is needed, however, to clarify which biological factors are most significant in creating and maintaining host related frequency differences at the Grant, Michigan site.

The geographic pattern of genetic differentiation between apple and hawthorn populations of *R. pomonella* appears to be complex. In an accompanying letter, McPheron *et al.*¹⁶ also document significant frequency differences at a sympatric site in Urbana, Illinois. Their results differ slightly from ours, however, in the pattern of loci and frequencies of alleles displaying inter-host differentiation. In addition, a recently completed analysis of geographic variation across the eastern United States (Feder *et al.*, manuscript in preparation) indicates that north-south allele frequency clines exist for both apple and hawthorn flies across their ranges for five of the six loci showing host-associated heterogeneity. Inter-host differences are therefore superimposed on latitudinal patterns of variation within the hawthorn and apple races and these clines help to explain the genetic differences between sites in Michigan and Illinois. The

results from the geographical survey raise the question of whether sympatric mechanisms alone account for the development of complete reproductive isolation between *R. pomonella* populations. We consequently believe that it is inappropriate to state definitively that hawthorn and apple races represent 'incipient' species. But the *R. pomonella* complex contains a number of sympatrically distributed sibling species specializing on different host plants which are inter-fertile in laboratory crosses, yet remain distinct in nature^{1,17}.

We acknowledge the critical comments of S. M. Williams, B. A. McPherson and S. H. Berlocher. Special thanks to S. Jamrog and A. Werthiemer for help in fly rearing.

Received 20 May; accepted 23 August 1988.

1. Bush, G. L. *The Taxonomy, Cytology and Evolution of the Genus Rhagoletis in North America* (Diptera, Tephritidae) (MCZ, Cambridge, Mass., 1966).
2. Bush, G. L. *Evolution* **23**, 237-251 (1969).
3. Bush, G. L. *A. Rev. Ecol. Syst.* **6**, 339-364 (1975).
4. Prokopy, R. J., Bennett, E. W. & Bush, G. L. *Can. Ent.* **103**, 1405-1409 (1971).
5. Prokopy, R. J., Bennett, E. W. & Bush, G. L. *Can. Ent.* **104**, 97-104 (1972).
6. Walsh, B. J. *Hort.* **2**, 338-343 (1867).
7. Cockerham, C. C. & Weir, B. S. *Genet. Res.* **30**, 121-147 (1977).
8. Laurie-Alhberg, C. C. & Weir, B. S. *Genetics* **92**, 1295-1314 (1979).
9. Berlocher, S. H. & Smith, D. C. *J. Hered.* **74**, 337-340 (1983).
10. Reissig, W. H. & Smith, D. C. *Ann. Ent. Soc. Am.* **71**, 155-159 (1978).
11. Felsenstein, J. *Heredity* **35**, 124-138 (1981).
12. Prokopy, R. J., Diehl, S. R. & Cooley, S. S. *Oecologia* **76**, 138-147 (1988).
13. Prokopy, R. J., Averill, A. L., Cooley, S. S. & Roitberg, C. A. *Science* **218**, 76-77 (1982).
14. Diehl, S. R. thesis, Univ. Texas at Austin (1984).
15. Smith, D. C. *Nature* **336**, 66-67 (1988).
16. McPherson, B. A., Smith, D. C. & Berlocher, S. H. *Nature* **336**, 64-66 (1988).
17. Berlocher, S. H. & Bush, G. L. *Syst. Zool.* **31**, 136-155 (1982).

Genetic differences between host races of *Rhagoletis pomonella*

Bruce A. McPherson*, D. Courtney Smith*
& Stewart H. Berlocher

Department of Entomology, University of Illinois, Urbana,
Illinois 61801, USA

Sympatric speciation by the formation of host races (parasite populations associated with different plant or animal hosts) has been the subject of great controversy. It has been difficult to demonstrate the existence of host races^{1,2}, much less to prove that host races are evolving toward species status. Genetic polymorphism attributable to association with different resources does occur³, but the phenomenon is far from ubiquitous in parasite populations. The apple maggot fly *Rhagoletis pomonella* uses a variety of host plants, and Bush^{4,5} has argued that it is a likely candidate for speciation by a sympatric mode. So far however there has been no direct evidence of any genetic differentiation between host-associated fly populations. We report significant differences in allele frequencies between fly populations reared from sympatric apple (*Malus pumila*) and hawthorn (*Crataegus mollis*) trees at a field site in Urbana, Illinois, in the United States.

Genetic heterogeneity was evident both in the five paired apple-hawthorn comparisons and in comparisons within a single host species over the five sites (Table 1; see Fig. 1 for the spatial arrangement of trees and the methods used). Allele frequencies at the enzymes cytosolic aconitase (*Acon-2*) and β -hydroxyacid dehydrogenase (*Had*) were consistently different between apple and hawthorn fly populations (Table 2). In each of the five paired apple-hawthorn comparisons, the frequency of *Acon-2*¹⁰⁰ was lower in the apple sample than in the corresponding hawthorn sample (Table 2). Conversely, frequencies of both *Acon-2*⁷⁵ and *Acon-2*¹⁰⁹ were higher in every

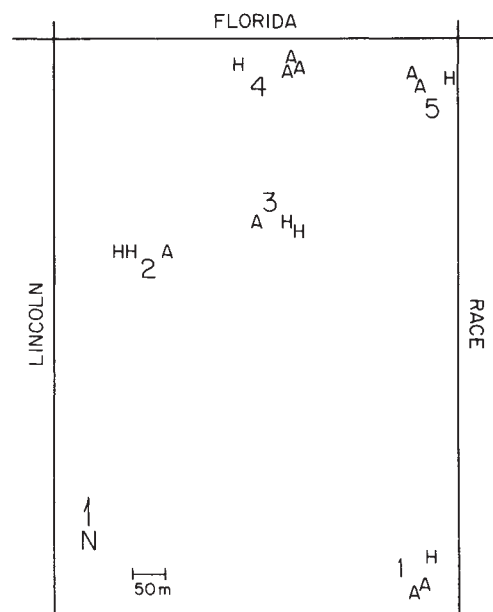


Fig. 1 Spatial arrangement of the five collecting sites in Urbana, Illinois. The symbol A refers to individual apple trees and the symbol H refers to individual hawthorn trees.

Methods. Infested fruit were collected from beneath five paired sets of apple (*M. pumila*) and hawthorn (*C. mollis*) trees. Apple collections were made in mid-August, 1985, and hawthorn collections followed in mid-September, 1985. Fruit were held under controlled conditions in the laboratory until larvae emerged and pupated in moist vermiculite. Pupae were then stored at 6 °C for six months to break diapause. Adult flies eclosed between one to three months following exposure to 24 °C at 20:4 light:dark. One hundred frozen adult flies from each host at each site were analysed by standard horizontal starch gel electrophoresis^{7,8,15-17} for 17 loci: aconitase-1, aconitase-2, adenylate kinase, alcohol dehydrogenase-1, aldolase, arginine kinase, aspartate aminotransferase-2, NADH-dependent diaphorase-2, fumarate hydratase, α -glycerophosphate dehydrogenase, β -hydroxyacid dehydrogenase, isocitrate dehydrogenase, mannose phosphate isomerase, phosphoglucose isomerase, phosphoglucomutase, trehalase, and triose phosphate isomerase. Mendelian inheritance of significantly variable loci has been demonstrated (refs 15 and 17; J. L. Feder, C. A. Chilcote and G. L. Bush, personal communication). Allele designations are by mobility relative to the most common allele at a locus (=100). χ^2 analysis of these loci revealed only five deviations from Hardy-Weinberg equilibrium out of 134 tests, a level expected from chance alone.

apple sample than in the corresponding paired hawthorn samples. *Had* allele frequencies displayed a similar consistency between the paired apple and hawthorn samples. The frequency of *Had*¹⁰⁰ was always greater in the apple sample than in the respective hawthorn sample, but the difference was not significant at site 5. Several other loci were heterogeneous for single site comparisons (Table 1) but did not display the consistent patterns seen in *Acon-2* and *Had*. Differences in host-associated loci between this study and that of Feder *et al.*⁶ reflect a strong latitudinal allele frequency cline at several enzyme genes in *R. pomonella* (J. L. Feder, C. A. Chilcote and G. L. Bush, personal communication and ref. 7).

Significant intrahost differentiation among both apple- and hawthorn-associated flies (Table 1) was of insufficient magnitude to obscure the patterns of interhost differences. Similar microgeographic heterogeneity has been demonstrated previously among flies from individual trees for another hawthorn-infesting *R. pomonella* population in Urbana, Illinois⁸. Cases of significant intrahost heterogeneity in our data were primarily

* Present addresses: Department of Entomology, Pesticide Research Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802, USA (B.A.M.); Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA (D.C.S.).

Table 1 *G*-tests for allele frequency heterogeneity at each of the five paired apple-hawthorn sites, across the five apple sites and across the five hawthorn sites

Comparison	<i>Acon-2</i>	<i>Aat-2</i>	<i>Dia-2</i>	<i>Had</i>	<i>Idh</i>	<i>Mpi</i>	<i>Pgi</i>	<i>Pgm</i>
Site 1	22.5 (4)*	5.3 (4)	1.5 (1)	19.6 (1)*	0.0 (1)	0.5 (1)	0.0 (1)	5.6 (2)
Site 2	32.9 (4)*	18.9 (4)*	18.8 (1)*	13.9 (1)*	0.3 (1)	0.3 (1)	4.0 (1)†	2.6 (2)
Site 3	31.8 (4)*	8.3 (4)	0.1 (1)	8.2 (1)*	4.5 (1)†	1.8 (1)	0.2 (1)	1.4 (2)
Site 4	36.8 (4)*	9.2 (4)	3.3 (1)	5.3 (1)†	0.5 (1)	9.1 (1)*	0.0 (1)	0.4 (2)
Site 5	26.0 (4)*	7.5 (4)	3.1 (1)	0.8 (1)	0.4 (1)	0.5 (1)	5.5 (2)	8.9 (2)†
All apples	56.6 (16)*	37.4 (16)*	6.9 (4)	3.3 (4)	10.9 (4)†	1.1 (4)	5.1 (4)	8.3 (8)
All hawthorns	24.5 (16)	26.2 (16)	14.5 (4)‡	12.1 (4)†	4.7 (4)	20.6 (4)*	1.4 (4)	28.9 (8)*

Allele classes were pooled where necessary to minimize classes with expected numbers less than five; degrees of freedom appear in parentheses.

* $P \leq 0.001$. † $P \leq 0.05$. ‡ $P \leq 0.01$.

Table 2 Allele frequencies of apple- and hawthorn-associated flies at two loci (*Acon-2* and *Had*) demonstrating consistent patterns of differentiation among hosts at each of the five sites

Locus/Allele	1A	1H	2A	2H	Site 3A	3H	4A	4H	5A	5H
<i>Acon-2</i>										
75	0.168	0.050	0.095	0.040	0.185	0.090	0.172	0.020	0.160	0.045
86	0.051	0.055	0.035	0.120	0.115	0.050	0.111	0.065	0.115	0.110
91	0.122	0.130	0.065	0.135	0.175	0.125	0.157	0.150	0.115	0.160
100	0.459	0.635	0.505	0.580	0.275	0.535	0.404	0.600	0.425	0.590
109	0.173	0.105	0.285	0.115	0.225	0.185	0.152	0.130	0.180	0.090
114	0.026	0.025	0.015	0.010	0.025	0.015	0.005	0.035	0.005	0.005
<i>Had</i>										
100	0.645	0.425	0.630	0.445	0.655	0.525	0.575	0.460	0.615	0.570
125	0.355	0.575	0.370	0.555	0.345	0.475	0.425	0.540	0.385	0.430

Sites are numbered with reference to Fig. 1 (A refers to the apple sample and H to the hawthorn sample at a site). Two rare alleles at *Acon-2* and four rare alleles at *Had* are pooled into common allele classes. See McPherson⁷ for complete data at these and other loci.

Table 3 F_{ST} analysis of the significantly polymorphic loci (95% criterion) in flies from both apples and hawthorn

Host	Locus										Average	±s.d.
	<i>Acon-2</i>	<i>Ak</i>	<i>Adh-1</i>	<i>Aat-2</i>	<i>Dia-2</i>	<i>Had</i>	<i>Idh</i>	<i>Mpi</i>	<i>Pgi</i>	<i>Pgm</i>		
Apple	0.015	0.003	0.000	0.005	0.004	0.001	0.007	0.000	0.000	0.001	0.005	±0.0024
Hawthorn	0.002	0.001	0.007	0.005	0.013	0.011	0.003	0.017	0.000	0.010	0.006	±0.002
Hierarchical	0.033	0.000	0.000	0.004	0.014	0.041	0.000	0.001	0.002	0.000	0.012	±0.0069

F_{ST} values, weighted average F_{ST} values and jack-knife variances were calculated by the methods of Weir and Cockerham¹⁴. The hierarchical F_{ST} calculations were also by the methods of Weir and Cockerham, treating the five apple collections and the five hawthorn collections as subdivisions within the total sample. There is no significant difference between the weighted average F_{ST} value from apple- and hawthorn-associated flies (t -test, $t = -1.053$, d.f. = 18, $P > 0.2$). The hierarchical F_{ST} is significantly greater than both the apple F_{ST} (approximate t -test, $t = -3.051$, d.f. = 9, $P < 0.02$) and the hawthorn F_{ST} (approximate t -test, $t = -2.641$, d.f. = 9, $p < 0.05$).

a result of single sites deviating from the average allele frequency for flies on that host, although no single site was responsible for most of the heterogeneous loci in either apple or hawthorn comparisons⁷. Also, no loci were significantly heterogeneous in both apple and hawthorn (Table 1). The significant heterogeneity of *Acon-2* at the apple sites and *Had* among the hawthorn samples was of particular concern, as these were the two loci displaying significant interhost differences. However, significance was due in each case to the contribution of a single site (site 3 for *Acon-2* in apples and site 5 for *Had* in hawthorns).

F_{ST} analysis, a method of quantifying genetic variation across sites, revealed an overall low level of intrahost differentiation (Table 3), although F_{ST} values varied by an order of magnitude among loci. Apple and hawthorn populations displayed similar levels of intrahost heterogeneity (Table 3). We used a hierarchical F_{ST} calculation to quantify the effect of host plant association on *R. pomonella* populations. The weighted hierarchical F_{ST} over 10 loci is significantly higher than the average F_{ST} for either apple or hawthorn sites alone (Table 3), indicating a greater role of host plant than spatial separation in fly population structure.

At least two hypotheses may be proposed to explain allele frequency differences between apple- and hawthorn-associated flies. First, the adult fly population may be panmictic, with females ovipositing in any available fruit. Interhost divergence could be generated by differential selection acting on genes closely linked to the enzyme loci in question (it is unlikely that selection is operating directly on at least some of the enzyme genes; *Acon-2* varies in both Illinois and Michigan⁶, but allele frequencies on the two hosts are completely reversed at these two locales). A second, although not mutually exclusive, possibility is that apple and hawthorn flies may have attained at least partial pre-mating reproductive isolation as a result of mating fidelity to their respective host plants⁵. Restricted gene flow has permitted differentiation, because of several conceivable causes⁹, at loci such as *Acon-2* and *Had*.

These two hypotheses can be tested. The most direct test involves sampling adult flies, preferably mating and ovipositing, from the two host plant species. From the first hypothesis, allele frequencies should be homogeneous, whereas the second hypothesis would predict frequency differences as reported here. Cross-rearing larvae from controlled matings in a split progeny

design would directly test whether selection by host-plant associated characters generates the heterogeneity observed in this study.

Pending completion of these tests, we do not know what factor is most responsible for host-associated genetic divergence at the Urbana, Illinois site. However, recent experiments demonstrate a genetic difference between apple and hawthorn flies with respect to seasonal emergence time. Flies from the same ten populations reported upon here were reared in apples under identical conditions and exposed to the same cold-treatment regimen to break diapause, and apple flies emerged consistently earlier than hawthorn flies (D.C.S., B.A.M. and S.H.B., unpublished results; similar results from sites near that of this study are reported by Smith¹⁰). This result is not only a further demonstration of genetic differences between two host-associated populations, but is also a biologically relevant difference. The native host of *R. pomonella* is hawthorn, and the fly began using apples only sometime after the introduction of that plant into North America¹¹. Apples and hawthorns exhibit different fruiting phenologies; apples are suitable for oviposition earlier in the season than hawthorn. This allochory is potentially important in serving to isolate sympatric populations on the two host plants¹², depending on factors such as the actual fidelity of adult flies to the host plant^{4,5} and overlap in fly emergence distributions. The effects of phenology on allele frequencies could be addressed by analysing flies from sympatric early- and late-fruited apples or hawthorns.

We have demonstrated that genetic differences exist between a sympatric apple and hawthorn *R. pomonella* population in Urbana, Illinois. Evidence is accumulating to suggest that host-associated genetic differences exist over the entire range of *R.*

pomonella, although the loci and patterns of allele frequency differences vary geographically^{6,7}. These differences are important irrespective of what causal agent is found to be responsible. Should the first hypothesis prove correct and adult flies are truly panmictic, the data are of great relevance to arguments over the role of environmental heterogeneity in maintaining genetic variation^{3,13}. If the second hypothesis is correct, as indicated by the different emergence phenology of apple and hawthorn flies, then our data provide strong support for the existence of host races on sympatric apple and hawthorn trees.

We acknowledge the critical comments of J. L. Feder, D. J. Futuyma, D. P. Pashley and W. S. Sheppard. This work was supported by the California Department of Food and Agriculture.

Received 19 May; accepted 4 September 1988.

1. Diehl, S. R. & Bush, G. L. *A. Rev. Ent.* **29**, 471-504 (1984).
2. Futuyma, D. J. & Mayer, G. C. *Syst. Zool.* **29**, 254-271 (1980).
3. Futuyma, D. J. & Peterson, S. C. *A. Rev. Ent.* **30**, 217-238 (1985).
4. Bush, G. L. in *Genetic Mechanisms of Speciation in Insects* (ed. White, M. J. D.) 3-23 (Australian and New Zealand Book Co., 1974).
5. Bush, G. L. in *Evolutionary Strategies of Parasitic Insects and Mites* (ed. Price, P. W.) 187-206 (Plenum, New York, 1975).
6. Feder, J. L., Chilcote, C. A. & Bush, G. L. *Nature* **336**, 61-64 (1988).
7. McPherson, B. A. thesis, Univ. Illinois (1987).
8. McPherson, B. A., Smith, D. C. & Berlocher, S. H. *Genetics* **119**, 445-451 (1988).
9. Slatkin, M. *A. Rev. ecol. Syst.* **16**, 393-430 (1985).
10. Smith, D. C. *Nature* **336**, 66-67 (1988).
11. Porter, B. A. *U.S.D.A. Tech. Bull.* No. 66 1-48 (1928).
12. Bush, G. L. *Bull. Mus. comp. Zool. Harv.* **134**, 431-562 (1966).
13. Hedrick, P. W. *A. Rev. Ecol. Syst.* **17**, 535-566 (1986).
14. Weir, B. S. & Cockerham, C. C. *Evolution* **38**, 1358-1370 (1984).
15. Berlocher, S. H. *J. Hered.* **71**, 63-67 (1980).
16. Berlocher, S. H. *A. ent. Soc. Amer.* **73**, 131-137 (1980).
17. Berlocher, S. H. & Smith, D. C. *J. Hered.* **74**, 337-340 (1983).

Heritable divergence of *Rhagoletis pomonella* host races by seasonal asynchrony

D. Courtney Smith

Department of Biology, 201 South Biology, University of Utah, Salt Lake City, Utah 84112, USA

Possible mechanisms for animal speciation in the absence of geographic isolation remain controversial. Univoltine flies of the *Rhagoletis pomonella* sibling species group of larval fruit parasites have been cited¹⁻⁵ as prime examples of sympatric speciation through divergence in host plant association. The recent emergence of several populations of *R. pomonella* associated with plants other than the fly's native hawthorn host has been attributed to the effects described by a model²⁻⁴ of sympatric host-race formation involving host recognition and survival genes; however, direct genetic evidence of biologically meaningful differentiation among putative host races has been lacking. In contrast, the potential importance of seasonal asynchrony (of flies adapted to host plants with different fruiting phenologies) in initial restriction of gene flow has been largely undervalued. Here I report heritable divergence in seasonal emergence of *R. pomonella* flies collected from sympatric populations on three host plant species in Illinois, in the United States.

The *R. pomonella* flies used in this study came from collections of hawthorn (*Crataegus mollis*), apple (*Malus pumila*) and dogwood (*Cornus florida*, a newly-determined host⁶) fruits made throughout Illinois. This study had two parts: the first determined if heritable differences in post-diapause eclosion time (PDET) existed among flies from these three host-associated populations (HAP); the second segregated observed differences into host-related and geographic effects. In each experiment,

measurements of PDET were determined for first-generation laboratory populations produced from separate mass crosses of the flies from the field-collected fruits rather than from those wild flies themselves. The resulting experimental flies were reared simultaneously under identical conditions to eliminate the possible effects of differences in nutrition or microhabitat.

The heritable nature of differences in PDET among these HAPs was evident from the results of laboratory crosses in which the offspring were artificially reared under identical circumstances and then exposed to seminatural conditions for temperature and photoperiod (Table 1), the two environmental factors with the greatest influence on eclosion time of temperate insects⁷. The reciprocal F₁ hybrid progeny were intermediate to the parental flies, demonstrating what is probably a mostly additive genetic effect. This seasonal life-history character may thus have the potential for rapid directional selection⁷ in response to adaptation to host species with different phenologies.

In Illinois, apple fruit generally matures during mid to late summer, hawthorn during early autumn and dogwood during mid to late autumn, an observation which is reflected in the order of mean eclosion times (Table 1) of the experimental fly populations. Although *R. pomonella* flies from the three hosts had significantly different mean PDET values, adaptation to climatic variation among geographic sites may cause some local fly populations to overlap more than others. Laboratory PDET of flies originating from apple, hawthorn or dogwood fruits collected from 13 locations throughout Illinois are shown in Table 2. Mean PDET values were significantly different for each HAP at every site containing more than one HAP. Geographic variation within a HAP was evident, with an 8-day maximum difference among sites. However, although these intrahost site differences had a significant effect ($F = 5.140$; 17 and 3391 d.f.; $P < 0.001$) on the overall variance, they provided less than 1% of the total. In contrast, PDET differences associated with host plant species ($F = 542.406$; 2 and 17 d.f.; $P < 0.001$) contributed over 77% of the total variance. Therefore, host plant phenology

Table 1 PDET of parental and reciprocal F_1 hybrid offspring of crosses of *R. pomonella* flies

Crosses $\text{♀} \times \text{♂}$	<i>n</i>	Mean $\pm$ standard error
A $\times$ A	217	40.0a $\pm$ 0.59
A $\times$ H	118	48.2b $\pm$ 0.74
H $\times$ A	160	47.3b $\pm$ 0.73
H $\times$ H	211	57.8c $\pm$ 0.87
H $\times$ D	222	65.1d $\pm$ 0.80
D $\times$ H	121	66.0d $\pm$ 0.95
D $\times$ D	157	76.7e $\pm$ 1.24

The naturally-infested fruits were brought to the laboratory and the ensuing fly pupae were stored in covered dishes of moist vermiculite at 4 °C for at least five months to terminate diapause. After removal from the cold, the dishes of pupae were placed under standard laboratory conditions for photoperiod (18:6 light:dark), temperature (24 $\pm$ 0.5 °C) and relative humidity (55 $\pm$ 5%). Eclosing flies were maintained in laboratory cages with a constant source of food (enzymatic yeast hydrolysate and brown sugar), water and cold-stored Red Delicious variety apples for oviposition and larval development. Laboratory-reared pupae were stored at 4 °C for six months. Following cold treatment, dishes of pupae were kept outdoors in a 35 cm³ screened field cage placed in continuous shade on bare ground before adult eclosion. PDET for each fly was measured from the day its pupa was removed from the cold. A, flies from apple; H, hawthorn; D, dogwood. Means followed by the same letter are not significantly different under Student-Newman-Keuls procedure at $P < 0.001$.

is important in limiting seasonal overlap of these flies. There were differences in the overall weighted means of almost four weeks between apple and hawthorn flies and over two weeks between hawthorn and dogwood flies. A similar temporal distance in nature would substantially restrict gene flow among HAPs and thereby facilitate divergent evolution.

Although these results clearly indicate genetic differences in PDET among apple, hawthorn and dogwood fly populations that satisfy accepted criteria^{8,9} for their designation as host races, it is uncertain how they reflect the extent of seasonal divergence in nature. The emergence of wild flies may be fine-tuned to coincide closely with fruit maturation if both parasite and host respond to the same microclimatic factors, providing even greater temporal divergence of HAPs than could be observed in the seminatural or laboratory conditions of this study. Equally however, intrapopulation variation in PDET, coupled with fly longevity^{10,11}, may allow considerable temporal overlap among some local HAPs. Previous reports of natural seasonal overlap between HAPs either fail to present supporting data^{2,4,12} or are confounded by a difference in geographic origin that is well beyond the normal flight range of the flies¹³.

Other features reported to separate these HAPs seem to represent the effects of restricted gene flow rather than potential causes. For example, not only has variation among HAPs for certain adult morphological characters^{1,2} not been correlated with restricted host selection or mating, but these size-related measurements may be confounded by possible nutritional

Table 2 Laboratory PDET of *R. pomonella* flies from different geographic collections in Illinois

Collection sites	Mean $\pm$ standard error (<i>n</i> flies)		
	Apple	Hawthorn	Dogwood
Mahomet	45.4 $\pm$ 0.72 (196)	<72.3 $\pm$ 0.87 (351)	
Champaign	47.1 $\pm$ 0.95 (149)		
Urbana-1	46.5 $\pm$ 0.64 (163)		
Urbana-2		74.3 $\pm$ 0.86 (318)	
Urbana-3	44.5 $\pm$ 0.35 (546)	<73.9 $\pm$ 0.73 (330)	
Urbana-4	47.4 $\pm$ 0.46 (274)		
Villa Grove	46.8 $\pm$ 0.91 (62)	<72.8 $\pm$ 0.93 (68)	
Charleston		74.1 $\pm$ 1.05 (49)	
Fox Ridge State Park			
Park	49.3 $\pm$ 0.79 (85)	<67.8 $\pm$ 1.96 (41)	<87.7 $\pm$ 1.14 (82)
Fairfield		75.0 $\pm$ 0.64 (196)	<87.4 $\pm$ 0.85 (135)
Harrisburg			85.2 $\pm$ 1.40 (55)
Carbondale		72.7 $\pm$ 1.53 (52)	<93.3 $\pm$ 0.56 (203)
Devil's Kitchen			
State Park			86.9 $\pm$ 1.90 (56)

The sites are listed in descending order from north to south in Illinois. The two furthest sites are about 300 km apart. Fly pupae were exposed to standard laboratory conditions (see Table 1). Means separated by the symbol (<) are significantly different using Student-Newman-Keuls procedure at $P < 0.001$.

differences among host species^{8,14}. In addition, significant allele frequency differences between HAPs revealed by enzyme electrophoresis¹⁵ may have arisen in one or more of several possible ways¹⁶ after segregation of flies on different host species. Moreover, those results do not even signify the existence of a reproductive barrier among populations. These HAPs still could represent a single panmictic population, with electrophoretic differences resulting from selection acting on the enzyme loci in question or on closely linked genes. The selective agents may involve differences in fruit chemistry or in other environmental factors (for example, temperature, water or plant nutrient stresses) related to the phenology or habitat of the host species. Therefore, one should not automatically attach evolutionary significance to population differentiation using a character whose biological relevance is uncertain.

Temporal asynchrony was first proposed¹ as the most likely mechanism for the initial divergence of new host races of these *Rhagoletis* flies. Because this seasonal divergence simultaneously restricts the availability of potential mates and hosts, PDET adaptation to their respective host species provides a parsimonious explanation for initial barriers to gene flow among fly populations. Because the relationship between host colonization and speciation is controversial^{8,14,17,18}, arguments concerning the likelihood of allopatric versus non-allopatric modes of speciation remain mostly conjectural, especially as little is known about the nature of gene flow even among fly populations on adjacent trees of a single host species¹⁹. Although the specific genetic nature of the developmental trait investigated here awaits elucidation, these results provide the first direct genetic evidence for a biologically meaningful character separating the host races of *R. pomonella* and also suggest that seasonal asynchrony is sufficient to both initiate and maintain restricted gene flow among fly populations from different host species.

I thank G. Edmunds and K. Paige for their comments.

Received 5 May; accepted 19 September 1988.

- Bush, G. L. *Bull. Mus. comp. Zool. Harv.* **134**, 431-562 (1966).
- Bush, G. L. *Evolution* **23**, 237-251 (1969).
- Bush, G. L. in *Genetic Mechanisms of Speciation in Insects* (ed. White, M. J. D.) 3-23 (Australian and New Zealand Book, Sydney, 1974).
- Bush, G. L. in *Evolutionary Strategies of Parasitic Insects and Mites* (ed. Price, P. W.) 187-206 (Plenum, London, 1975).
- Bush, G. L. & Diehl, S. R. in *Insect-Plant Relationships* (eds Visser, J. H. & Minks, A. K.) 297-305 (Pudoc, Wageningen, 1982).
- Smith, D. C. *J.N.Y. ent. Soc.* (in the press).
- Tauber, M. J., Tauber, C. A. & Masaki, S. *Seasonal Adaptations of Insects* (Oxford, New York, 1986).

- Jaenike, J. *Am. Nat.* **117**, 830-834 (1981).
- Diehl, S. R. & Bush, G. L. *A. Rev. Ent.* **29**, 471-504 (1984).
- Reissig, W. H. *Can. Ent.* **109**, 631-634 (1977).
- Opp, S. B. & Prokopy, R. J. *Fla. Ent.* **70**, 449-457 (1987).
- Pickett, A. D. & Neary, M. E. *Scient. Agric.* **20**, 551-556 (1940).
- Reissig, W. H. & Smith, D. C. *A. ent. Soc. Am.* **71**, 155-159 (1978).
- Futuyma, D. J. & Mayer, G. C. *Syst. Zool.* **29**, 254-271 (1980).
- Smith, D. C. dissertation, Univ. Illinois (1986).
- Slatkin, M. A. *Rev. Ecol. Syst.* **16**, 393-430 (1985).
- Paterson, H. E. H. *S. Afr. J. Sci.* **77**, 113-119 (1981).
- Futuyma, D. J. & Peterson, S. C. *A. Rev. Ent.* **30**, 217-238 (1985).
- McPherson, B. A., Smith, D. C. & Berlocher, S. H. *Genetics* **119**, 445-451 (1988).

NMDA receptors activate the arachidonic acid cascade system in striatal neurons

A. Dumuis, M. Sebben, L. Haynes,
J.-P. Pin & J. Bockaert

Centre CNRS-INSERM de Pharmacologie-Endocrinologie,
Rue de la Cardonille, 34094 Montpellier Cedex 2, France

Receptors for excitatory amino-acid transmitters on nerve cells fall into two main categories associated with non-selective cationic channels, the NMDA (*N*-methyl-D-aspartate) and non-NMDA (kainate and quisqualate) receptors^{1,2}. Special properties of NMDA receptors such as their voltage-dependent blockade by Mg^{2+} (refs 3, 4) and their permeability to Na^+ , K^+ as well as to Ca^{2+} (refs 5, 6), have led to the suggestion that these receptors are important in plasticity during development and learning. They have been implicated in long-term potentiation⁷ (LTP), a model for the study of the cellular mechanisms of learning⁷⁻¹¹. We report here that glutamate and NMDA, acting at typical NMDA receptors, stimulate the release of arachidonic acid (as well as 11- and 12-hydroxyeicosatetraenoic acids from striatal neurons probably by stimulation of a Ca^{2+} -dependent phospholipase A_2). Kainate and quisqualate, as well as K^+ -induced depolarization were ineffective. Our results provide direct evidence in favour of the hypothesis^{12,13}, that arachidonic acid derivatives, produced by activation of the postsynaptic cell, could be messengers that cross the synaptic cleft to modify the presynaptic functions known to be altered during LTP⁹. In addition, we suggest that NMDA receptors are the postsynaptic receptors which trigger the synthesis of these putative transynaptic messengers.

Primary cultures of striatal neurons generated from 14–15-day-old mouse embryos were prepared as previously described¹⁴ and maintained for six days in serum-free medium. At this developmental stage, differentiated synapses were not observed¹⁴. Neurons in six well plates were incubated overnight with [3H] arachidonic acid. Ninety per cent of added label was incorporated into striatal neurons. Before the start of the experiment, the cells were washed with a Krebs bicarbonate medium. After the final wash, the experimental agents were added in 2 ml Krebs bicarbonate buffer. Excitatory amino acid (EAA)-induced [3H] arachidonic acid metabolites ([3H]AAM) release was measured at the times indicated by determining the radioactivity in the media. Basal [3H]AAM release increased for 10 min and reached a plateau after 15 min (data not shown). Glutamate (30 μM) rapidly stimulated basal [3H]AAM release, and by 10 min 1–2.5% of the [3H]arachidonic acid incorporated into the neurons was released in the presence of glutamate. After 20 min we deduced that the effect of glutamate tended to decrease, as a plateau in basal and stimulated release was reached (data not shown).

To determine the nature of the EAA receptor type(s) involved in this response, dose-response curves for glutamate, NMDA, kainate and quisqualate were first performed in the absence of Mg^{2+} , a condition necessary to yield full activity of NMDA receptors^{3,4}. Glutamate stimulated [3H]AAM release by a factor of two or three. Half-maximal stimulation (EC_{50}) was obtained with 8 $\mu M \pm 1.4$ ($n = 6$). The maximal effect of NMDA ($EC_{50} = 25 \mu M \pm 6$, $n = 6$) was always slightly lower than the maximal glutamate response (Fig. 1). Kainate and quisqualate were totally inactive (Fig. 1a). The low Mg^{2+} concentration present in the Krebs bicarbonate buffer (1.2 mM), had very little effect on the glutamate dose-response curve (Fig. 1b) which was similar to that obtained in the absence of Mg^{2+} (Fig. 1a), whereas the maximal NMDA effect was reduced to 45% of the maximal glutamate response (Fig. 1b). Kainate and quisqualate were still inactive in the presence of 1.2 mM Mg^{2+} . High concentrations of Mg^{2+} completely blocked both the glutamate and NMDA

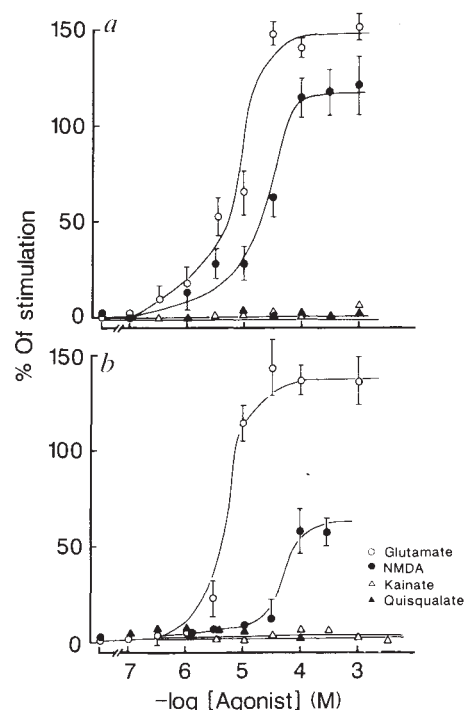


Fig. 1 Dose-dependent glutamate, NMDA, kainate and quisqualate effects on [3H]arachidonic acid and/or its metabolites release from striatal neurons. *a*, Krebs bicarbonate medium containing $MgSO_4$ (1.26 mM). *b*, $MgSO_4$ was omitted. The results are expressed as a percentage of stimulation over the basal activity. Results are the means $\pm$ s.e.m. of six separate experiments, each performed in triplicate. The basal and maximal stimulated release of [3H]arachidonic acid and/or metabolites were respectively $0.9 \pm 0.15\%$ and $2.3 \pm 0.2\%$ of the total radioactivity in the cell. Cells were extracted with 2 ml 0.2 N NaOH and 0.5% Triton X-100. **Methods.** Striatal neurons, after six days *in vitro*, were incubated overnight with 0.25 $\mu Ci ml^{-1}$ [3H]arachidonic acid (0.5 μCi per well) (specific activity 166 Ci $mmol^{-1}$) (CEA Saclay, France). Before the start of the experiments, neurons were washed five times at 5-min intervals with 2 ml per well of Krebs bicarbonate buffer (3.5 mM KCl, 0.75 mM $CaCl_2$, 124 mM NaCl, 1.25 mM KH_2PO_4 and 26.3 mM $NaHCO_3$, pH 7.4), 0.05% of albumin bovine, fatty acid-free was added and the buffer was saturated with 95% $O_2/5\%$ CO_2 . Neurons were then exposed for 15 min to the stimulation medium containing glutamate, kainate, NMDA or quisqualate, at the indicated concentration.

responses. However, Mg^{2+} was more potent in blocking the NMDA than the glutamate-induced [3H]AAM release (50% inhibitory concentration, $IC_{50} = 0.3$ and 2 mM, respectively, data not shown).

The selective competitive antagonist of NMDA¹⁵ receptors, 2-amino-5-phosphonopentanoate (APV), inhibited glutamate ($10^{-4} M$) and NMDA ($10^{-4} M$) responses in a dose-dependent manner ($IC_{50} = 10 \mu M \pm 2.1$, $n = 4$ and $12 \mu M \pm 3.8$, $n = 4$, respectively) (Fig. 2a, b). Furthermore, phencyclidine (PCP), a non-competitive antagonist¹⁶ of the NMDA channel, was also able to block the glutamate and NMDA responses with an IC_{50} of 50 ± 6.2 and $800 nM \pm 210$ ($n = 4$), respectively (Fig. 2c, d). As expected, glutamate stimulation of [3H]AAM release was inhibited competitively by APV and non-competitively by PCP (data not shown).

The glutamate induced increase in [3H]AAM release was dependent on external Ca^{2+} . When Ca^{2+} was omitted from the buffer, and when this omission was combined with the addition of 1 mM EGTA, glutamate did not stimulate [3H]AAM release (Table 1). Ionomycin (10 μM), a synthetic Ca^{2+} ionophore, greatly stimulated [3H]AAM release. In contrast, no effect on release was observed when the K^+ concentration was increased up to 56 mM (Table 1).

Table 1 Effect of K^+ , Ca^{2+} , ionomycin and mepacrine on basal and glutamate-stimulated release of [3H]arachidonic acid and/or its metabolites

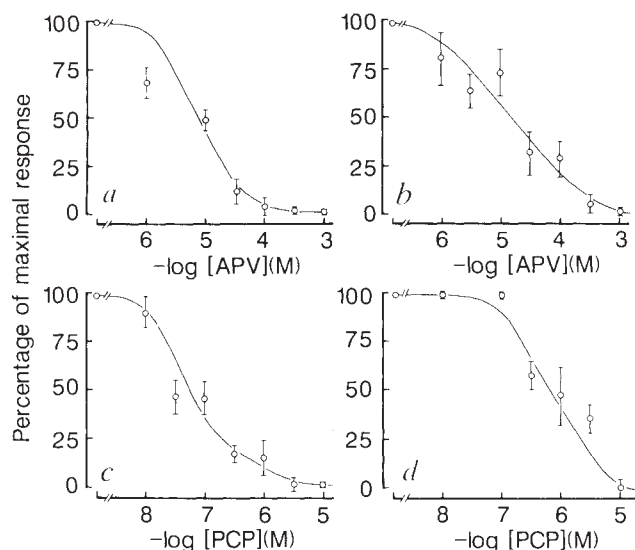
	Basal release (% Cell content)	Glutamate (3×10^{-5} M)
Control	0.73 ± 0.07 ($n=5$)	ND
K^+ (10 mM)	0.88 ± 0.10 ($n=5$)	ND
K^+ (25 mM)	0.85 ± 0.14 ($n=5$)	ND
K^+ (56 mM)	0.78 ± 0.10 ($n=5$)	ND
Control	0.63 ± 0.03 ($n=4$)	1.63 ± 0.10 ($n=4$)
Without [Ca^{2+}]	0.93 ± 0.07 ($n=4$)	0.96 ± 0.04 ($n=4$)
Without [Ca^{2+}] + 1mM EGTA	0.95 ± 0.06 ($n=4$)	0.92 ± 0.09 ($n=4$)
Control	0.72 ± 0.05 ($n=5$)	ND
Ionomycin (10^{-5} M)	5.36 ± 0.70 ($n=5$)	ND
Ionomycin (10^{-5} M) + mepacrine (10^{-4} M)	1.08 ± 0.16 ($n=5$)	ND
Control	1.05 ± 0.10 ($n=4$)	2.30 ± 0.24 ($n=4$)
Mepacrine (10^{-5} M)	0.96 ± 0.20 ($n=4$)	1.50 ± 0.21 ($n=4$)
Mepacrine (3×10^{-5} M)	1.05 ± 0.18 ($n=4$)	1.06 ± 0.17 ($n=4$)
Mepacrine (10^{-4} M)	0.90 ± 0.12 ($n=4$)	0.95 ± 0.05 ($n=4$)

Neuronal cultures were incubated overnight with $0.25 \mu Ci ml^{-1}$ of [3H]arachidonic acid. After washing, neurons were exposed to the stimulation medium for 15 min. In experiments to investigate the effect of Ca^{2+} , Ca^{2+} was omitted from the washing Krebs bicarbonate buffer; EGTA was added only when neurons were exposed to the stimulation medium. To investigate the effect of mepacrine, cells were exposed to mepacrine for 30 min before the experiments and it was present in the buffer during washing and the experiments. Results are the means $\pm$ s.e.m. of at least four independent experiments performed in triplicate. ND, not determined.

One explanation for our observations is the activation of phospholipase A_2 (PLA $_2$) by NMDA and ionomycin and the concomitant increase in arachidonic acid release. To test this hypothesis, we pre-exposed the cultures to mepacrine, an inhibitor of PLA $_2$ ¹⁷, before applying the agonists. Low concentrations of mepacrine (30–100 μM) completely abolished glutamate and ionomycin effects (Table 1).

To characterize the [3H]AAM released in the medium, striatal neurons were labelled with a large amount of [3H]arachidonic acid ($10 \mu Ci ml^{-1}$) for 24 h. The tritiated materials released from cells under basal and 100 μM glutamate-stimulated conditions were then analysed by HPLC¹⁸ (Fig. 3). Most of the radioactivity released under both conditions (78% under basal and 93% under glutamate stimulation) had the same retention time as authentic arachidonic acid (Fig. 3a). The amount of [3H]arachidonic acid released from the cells was greatly increased under glutamate stimulation (Fig. 3a). In addition to arachidonic acid, other small peaks were also increased by glutamate (Fig. 3b). Amongst them were two peaks which possess the same retention time as authentic 11- and 12-hydroxy-eicosatetraenoic acids (HETEs) (Fig. 3b).

These results indicate that, in striatal neurons labelled with [3H]arachidonic acid, [3H]AAM released by excitatory amino acids corresponds mainly to arachidonic acid, and that this effect is mediated to a large extent by the activation of a typical NMDA receptor. Both glutamate and NMDA stimulations were suppressed by Mg^{2+} , a divalent cation known to enter the open NMDA channel and block it^{3,4}. However, it is interesting that at low Mg^{2+} concentrations (1.2 mM) NMDA responses were more affected than glutamate responses (Fig. 1a, b, data not shown). One possible explanation is that glutamate, but not NMDA, depolarized striatal neurons by stimulating non-NMDA receptors (kainate/quisqualate) known to be present on these cells^{5,19}. Such a depolarization might overcome the blockade of NMDA channels by Mg^{2+} . The marked sensitivity of the NMDA response to external Mg^{2+} concentration

**Fig. 2** Inhibition by APV and PCP of glutamate and NMDA induced [3H]arachidonic acid and/or metabolite release. Results are given as a percentage of the maximal response (in the absence of antagonists) and are the means $\pm$ s.e.m. of triplicate determinations of four independent experiments.

Methods. Striatal neurons, after six days *in vitro*, were incubated overnight with [3H]arachidonic acid, washed five times and exposed to 100 μM glutamate for 15 min in the presence of varying concentrations of APV (racemic mixture) and PCP (a and c respectively) or to 100 μM NMDA for 15 min in the presence of varying concentrations of APV and PCP (b and d respectively).

described here is reflected by the fact that a reduced but significant NMDA action was observed at physiological Mg^{2+} concentrations.

Our proposal that glutamate and NMDA responses are mediated by NMDA receptors is also supported by their blockade by APV, a specific competitive antagonist of the NMDA receptor. Although the IC_{50} s of APV in inhibiting glutamate and NMDA-evoked [3H]arachidonic acid release were similar (about 10 μM), the dose-inhibition curves do not have the same shape (Fig. 2a, b). This is probably because the NMDA receptor is a complex structure with allosteric properties². This is suggested in our studies by the very steep dose-response curves (Fig. 1a).

PCP and other dissociative anaesthetic drugs such as ketamine, are believed to bind inside the open channel of the NMDA receptor²⁰. PCP was about 10 times more potent in blocking the glutamate response (determined in the presence of 1.2 mM Mg^{2+}) than the NMDA response (determined in the absence of Mg^{2+}) (Fig. 2c, d).

The potent inhibition by mepacrine, a PLA $_2$ inhibitor, of the glutamate- and ionomycin-stimulated [3H]arachidonic acid release (Table 1), indicates that both drugs activate PLA $_2$. It is likely that the well-described Ca^{2+} influx through NMDA receptor channels^{3,6} is the key element in PLA $_2$ activation by EAA receptors. The involvement of a G protein in the coupling between NMDA receptors and PLA $_2$ similar to the one described in thyroid cells between α_1 -adrenergic receptors and PLA $_2$ (ref. 21) is unlikely but cannot be totally excluded because in the latter system the coupling was also dependent on external Ca^{2+} (ref. 21). We did not observe any blocking effect of pertussis toxin on glutamate-induced [3H]arachidonic acid release (data not shown).

In this model system, kainate and quisqualate were totally ineffective in increasing [3H]arachidonic acid release (Fig. 1). These results can easily be explained by the previous observation that non-NMDA receptors trigger less neuronal Ca^{2+} entry than

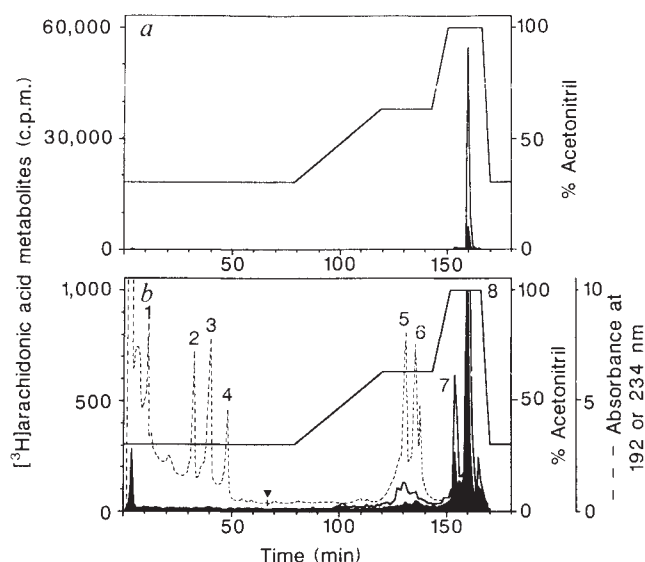


Fig. 3 HPLC analysis of the $[^3\text{H}]$ metabolites released from striatal neurons under basal (■) or 100 μM glutamate-stimulated (□) conditions. *b*, as *a* except that the c.p.m. values are cut off over 1,000 c.p.m. Total c.p.m. in the peak corresponding to arachidonic acid were 6,515 c.p.m. under basal conditions and 64,951 c.p.m. under glutamate stimulation. The elution of unlabelled prostaglandin standards was followed by measuring the absorbance at 192 nm (Merck spectrophotometer UV L 4000) during the first 68 min of the run (▼). (Dotted line in *b*, arbitrary units.) The HETE standards were then detected at 234 nm. 1, 6-keto-prostaglandin $\text{F}_{1\alpha}$; 2, prostaglandin $\text{F}_{2\alpha}$; 3, prostaglandin E_2 ; 4, prostaglandin D_2 ; 5, 11-HETE; 6, 12-HETE; 7, 5-HETE; 8, arachidonic acid.

Methods. Striatal neurons grown in 100 mm diameter Petri dishes (15×10^6 cells per dish) and labelled for 24 h with $[^3\text{H}]$ arachidonic acid ($10 \mu\text{Ci ml}^{-1}$) were treated as described in the Fig. 1 legend. After five washes with 10 ml of medium, neurons were exposed for 15 min to 3 ml of normal medium (basal) or medium containing 100 μM glutamate. The removed media were then subjected to HPLC analysis (Beckman gradient system) of the $[^3\text{H}]$ metabolites as previously described¹⁸. Briefly, 200 μl of medium containing the released $[^3\text{H}]$ compounds in which 5 μg each of unlabelled standards were added, were injected onto an Ultrasphere-ODS guard column connected to an Ultrasphere-ODS column (5 μm particles, 0.46×25 cm, Beckman). Elution was at room temperature, at a flow rate of 1 ml min^{-1} , with a mixture of H_2O (equilibrated at $\text{pH} = 3.5$ with acetic acid) and pure acetonitril (Rathburn). Evolution of the percentage (v/v) of acetonitril in eluent is shown. Fractions were taken every 1 min (1 ml) and counted in minivials containing 4 ml scintillation liquid (Amersham, ACS II).

NMDA receptors^{5,6}. Furthermore, Ca^{2+} entry triggered by kainate and quisqualate is due to activation of voltage-dependent Ca^{2+} channels⁵. We observed that since activation of these channels by K^+ does not trigger $[^3\text{H}]$ arachidonic acid release, the absence of kainate and quisqualate effects can be expected. In addition, the response was obtained in cultures in which no functional synapses and therefore no Ca^{2+} -dependent release of neurotransmitters was observed¹⁴, providing evidence that EAA-stimulated arachidonic release is direct rather than indirect.

The roles of arachidonic acid and/or its derivatives in the physiology of NMDA receptors remain to be determined. There have been two studies which suggested a role of PLA_2 in LTP. One reported that mepacrine inhibits the time course of a persistent LTP response and that this inhibition was reversed by *cis* unsaturated fatty acid²². The involvement of NMDA receptors in LTP induction is now well recognized⁷. Very recently, William and Bliss demonstrated¹³ that the lipoxygenase inhibitor, nordihydroguaiaretic acid, greatly reduces the induction of LTP in hippocampus. We propose that NMDA receptors

trigger biochemical events of LTP induction by an activation of the arachidonic acid cascade, similar to the one described here. The exact role of arachidonic and/or its metabolites (HETEs) in the synapse remains to be determined. It is possible that these compounds act postsynaptically to stimulate one form of protein kinase C^{23} or to induce genetic events²⁰. Alternatively, they may be released from the postsynaptic cell to act on the presynaptic terminal to trigger actions such as the increase in glutamate release²⁴ or inhibition of glutamate re-uptake²⁵. The latter hypothesis is consistent with the evidence of the existence of presynaptic modification during LTP¹¹.

Whatever the final answer, the results presented here further support the concept that EAA receptor actions are mediated both by membrane permeability changes and by production of chemical messengers²⁶⁻²⁸.

We thank Dr M. Lafon-Cazal for stimulating discussions and for her participation in determining the $[^3\text{H}]$ AAM and Mrs Angie Turner-Madeuf as well as Mrs M. Passama for their help in preparing this manuscript. Support was provided in part by the CNRS, INSERM, DRET and by grants from BAYER TROPONWERKE (France and FRG). L. Haynes was supported by a travelling fellowship from the Wellcome Trust.

Received 26 August; accepted 26 September 1988.

- Watkins, J. C. & Evans, R. H. *A. Rev. Pharmac. Tox.* **21**, 165-204 (1981).
- Mayer, M. L. & Westbrook, G. L. *Prog. Neurobiol.* **28**, 197-276 (1987).
- Nowak, L., Bregestovski, P., Ascher, P., Herbet, A. & Prochiantz, A. *Nature* **307**, 462-465 (1984).
- Mayer, M. L. & Westbrook, G. L. *J. Physiol., Lond.* **354**, 29-53 (1984).
- Mayer, M. L., MacDermott, A. B., Westbrook, G. L., Smith, S. J. & Barker, J. L. *J. Neurosci.* **7**, 3230-3244 (1987).
- Murphy, S. N., Thayer, S. A. & Miller, R. J. *J. Neurosci.* **7**, 4145-4158 (1987).
- Collingridge, G. L., Kehl, S. J. & McLennan, H. J. *Physiol., Lond.* **334**, 33-46 (1983).
- Bliss, T. V. P. & Lomo, T. *J. Physiol., Lond.* **232**, 331-356 (1973).
- Bliss, T. V. P. & Gardner-Medwin, A. R. *J. Physiol., Lond.* **232**, 357-374 (1973).
- Andersen, P., Sunberg, S. H., Sveen, O. & Wigström, H. *Nature* **266**, 736-737 (1977).
- Dolphin, A. C., Errington, M. L. & Bliss, T. V. P. *Nature* **297**, 496-498 (1982).
- Piomelli, D. *et al. Nature* **328**, 38-43 (1987).
- Williams, J. H. & Bliss, T. V. P. *Neurosci. Lett.* **88**, 81-85 (1988).
- Weiss, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 2238-2242 (1986).
- Watkins, J. C. & Olverman, H. J. *Trends Neurosci.* **10**, 265-272.
- Honey, C. R., Miljkovic, Z. & MacDonald, J. F. *Neurosci. Lett.* **61**, 135-139 (1985).
- Irvine, R. F. *Biochem. J.* **204**, 3-16 (1982).
- Van Rollins, M., Aveland, M. I., Sprecher, H. W. & Horrocks, L. A. *Meth. Enzym.* **86**, 518-530 (1982).
- Pin, J. P., Van Vliet, B. J. & Bockaert, J. *Neurosci. Lett.* **87**, 87-92 (1988).
- MacDonald, J. F., Miljkovic, Z. & Pennefather, P. J. *Neurophysiol.* **58**, 251-266 (1987).
- Burch, R. M., Luini, A. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7201-7205 (1986).
- Linden, D. J., Sheu, F. S., Murakami, K. & Routtenberg, A. *J. Neurosci.* **7**, 3783-3792 (1987).
- Sekiguchi, K., Tsukuda, M., Ogita, K., Kikkawa, U. & Nishizuka, Y. *Biochem. biophys. Res. Commun.* **145**, 797-802 (1987).
- Rhoads, D. E., Osburn, L. D., Peterson, N. A. & Raghupathy, E. *J. Neurochem.* **41**, 531-537 (1983).
- Chan, P. H., Kerlan, R. & Fishman, R. A. *J. Neurochem.* **40**, 309-316 (1983).
- Sladeczek, F., Pin, J. P., Récasens, M., Bockaert, J. & Weiss, S. *Nature* **317**, 717-719 (1985).
- Sugiyama, H., Ito, I. & Hirono, C. *Nature* **325**, 531-533 (1987).
- Nicoletti, F., Iadarola, M. J., Wroblewski, J. P. & Costa, E. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1931-1935 (1986).

B-cell memory is short-lived in the absence of antigen

David Gray & Helena Skarvall

Basel Institute for Immunology, Grenzacherstrasse 487, Postfach CH-4005 Basel, Switzerland

Primary encounter with antigen stimulates specific B cells not only to differentiate into cells that produce antibody at a high rate (plasma cells), but also to give rise to populations of memory cells. These cells have many characteristics that differ from virgin B cells, including their lifespan¹. When re-exposed to antigen, memory cells generate secondary IgG responses that are enhanced in rate, titre and affinity. At present they are considered as small resting lymphocytes which survive for long periods in a quiescent state between each antigen encounter². However, the fact that an individual may continue to make an antibody response for many

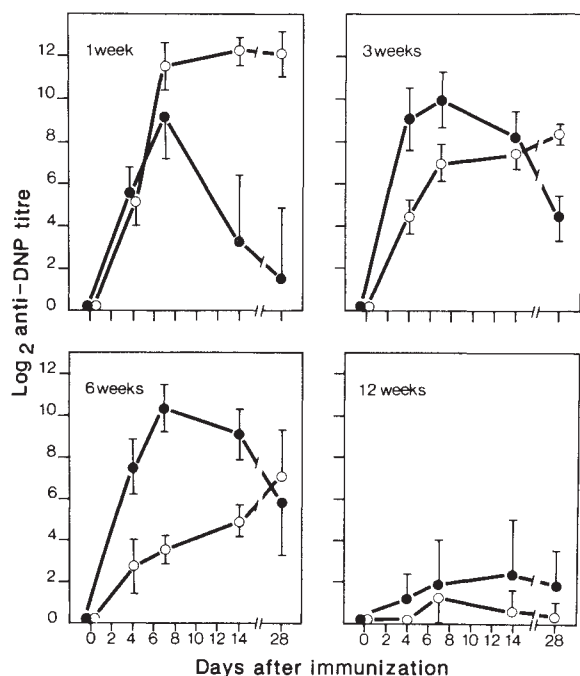


Fig. 1 The decay of memory B-cell populations in adoptive hosts in the absence of antigen. The figure shows the serum anti-DNP responses elicited by i.v. injection of 50 μ g soluble DNP-MSH (MSH, Maia squinada haemocyanin) from host, virgin cells (●) or from donor memory cells (○) in adoptive hosts at 1, 3, 6 and 12 weeks after transfer. The mean and standard deviation of the responses in 5–7 chimaeric animals is shown.

Methods. Congenic strains of rats differing only in their *k*-allotype were used^{3,24}. The κ 1b donors were primed (50 μ g alum-precipitated DNP-MSH injected i.p.) and boosted one month later (50 μ g DNP-MSH injected i.p.) and then used for thoracic-duct cannulation two to three months afterwards. Transfer of thoracic duct cells beforehand can lead to antigen transfer, as evidenced by donor anti-DNP titres without active immunization. Recipient κ 1a rats were prepared one day before transfer by irradiation with 5.5 Gy (from a ¹³⁷Cs γ -source; Atomic Energy Canada Ltd). After this dose, B cells from the host bone marrow begin to repopulate peripheral lymphoid tissues within one week. Each rat received 10⁸ donor-derived thoracic duct lymphocytes (TDL). These chimaeras were then immunized i.v. with 50 μ g soluble DNP-MSH at 1, 3, 6 or 12 weeks following transfer. All chimaeras were producing no donor or host anti-DNP antibody before immunization, indicating that no antigen was transferred. Serum responses of donor and host origin occurring over 28 days were measured in a radioimmunoassay using MARK-3 and FH8 monoclonal antibodies respectively. The antibodies and assay are described in detail elsewhere^{3,7}. MSH was supplied by the late Professor J. H. Humphrey, Royal Postgraduate Medical School, Hammersmith Hospital, London and dinitrophenylated to a substitution ratio of 172 molecules of DNP/10⁶ MSH, as described elsewhere²⁵.

months following a single injection of antigen^{3,4} is often overlooked. This continued antibody production is probably due to repeated stimulation of antigen-specific B cells and raises the question of whether memory B-cell clones require antigen for their maintenance. Here we show that they do, and that following transfer, in the absence of antigen, memory B-cell populations are lost from the adoptive host after 10–12 weeks.

To estimate the half life of memory B cells in the absence of antigen, thoracic duct lymphocytes from κ 1b rats primed with 2,4-dinitrophenyl-haemocyanin (DNP-MSH) were transferred into sublethally irradiated κ 1a congenic rats using procedures to minimize possible antigen transfer to the adoptive host. These cells were left in adoptive hosts for varying periods before immunization. We transferred both primed B and T cells so that we could compare the kinetics of donor memory and host virgin

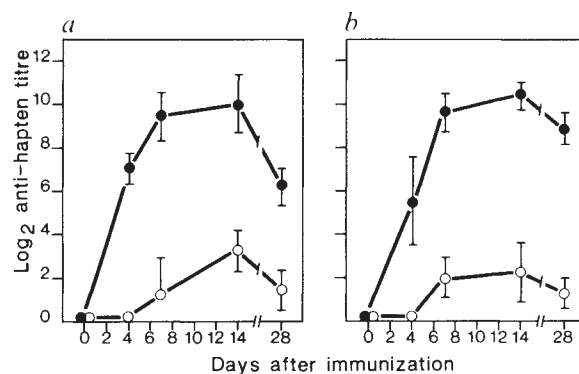


Fig. 2 Sixteen weeks after transfer, donor memory (○) responses are absent after immunization with alum-precipitated DNP-MSH plus 10⁹ *B. pertussis*, whereas the host response is restored. The figure shows the response to this form of DNP-MSH by the seven chimaeras already analysed in Fig. 1 (12 weeks) that gave no response to soluble DNP-MSH. The titre and kinetics of the anti-DNP response (a) are compared with the primary response to oxazolone-MSH (b). 2-phenyl-oxazolone was used to label MSH as described²⁶. See legend to Fig. 1 for other details.

responses when T-cell help was not a limiting factor. Figure 1 shows the host (virgin) and donor (memory) responses to an i.v. injection of soluble DNP-MSH in chimaeras immunized 1, 3, 6 or 12 weeks after transfer of cells. One week after transfer the memory (donor) response is rapid and dominates within seven days, but it should be noted that the host (virgin) response is equally rapid in these chimaeras. The host response falls in the second week, presumably as a result of competition with higher-affinity donor clones for limiting antigen^{3,5,6}.

After three weeks there has already been a decay in the memory-cell population, and by six weeks this is more marked. But after 12 weeks no donor (memory) response can be elicited. Because 12-week chimaeras showed very poor host (virgin) responses to soluble DNP-MSH, they were immunized a second time with antigen plus adjuvant. Figure 2 shows that alum-precipitated antigen plus *Bordetella pertussis* elicits a host response, whereas the donor (memory) response remains absent. The loss of the donor response is unlikely to be due to suppression or rejection of donor allotype cells because they survive for long periods following immunization (Fig. 3a), and indeed survive for at least six weeks in non-immunized chimaeras (Fig. 1c). We have shown previously in similar chimaeras that loss of specific donor responses is not due to allotype suppression or rejection because primary responses to unrelated antigens can be elicited^{7,8}.

The results demonstrate that, in the absence of antigenic stimulation, memory B-cell clones have a half life of 2–3 weeks. This estimate correlates well with an estimate of the lifespan of B cells in the recirculating pool of rats⁹. Askonas *et al.*¹⁰ have previously shown that dominant clones of memory B cells were sometimes lost when antigen was not given at the time of adoptive transfer, but they did not extend their analysis sufficiently to estimate the half life. These and other investigations of memory B-cell half lives (estimates of ~100 days) have been limited by less-than-rigorous attempts to prevent transfer of antigen with cells^{11,12}, and indeed Celada¹² reports residual antibody production without further challenge.

It is clear that memory B-cell clones may be very long-lived in the presence of antigen. The long-term response of primed donor populations transferred into adoptive hosts with soluble DNP-MSH is shown in Fig. 3a. Twenty weeks after transfer these chimaeras are still making serum antibody of donor allotype which can be boosted further. The maintenance of inert antigens (such as hapten-protein-conjugates) in lymphoid tissues is almost certainly in the form of antigen-antibody complexes on the surface of follicular dendritic cells^{4,13,14}, where

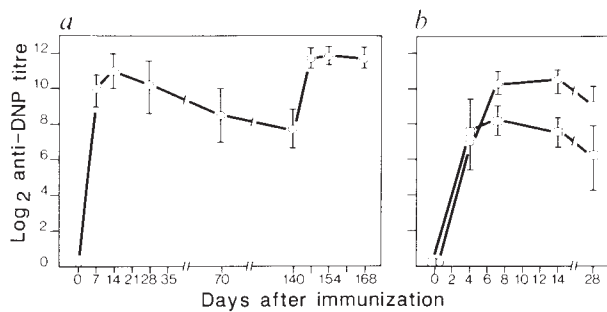


Fig. 3 *a*, Long-term donor response to DNP-MSH in chimaeras that received 5 μ g soluble DNP-MSH at the same time as the transfer of primed donor cells. A donor, memory response can be elicited in these animals at 20 weeks after transfer. Means and standard deviations of titres from five animals are shown. *b*, Transfer of donor, memory responses to DNP-MSH by cells having high (○) or low (□) buoyant density. Adoptive responses are transferred by both small, resting and large, activated B cells.

Methods. TDL from primed donors was depleted of T cells using magnetic beads (Dynabeads, Dynal AS, Oslo, Norway) coated with a pan-T monoclonal antibody W3/13 (ref. 27). Coated beads were mixed with TDL at a ratio of 40:1. The tube was placed on ice for 30 min and resuspended every 5 min. This procedure produced populations containing fewer than 1% T cells, as assessed by staining with fluorescein isothiocyanate-labelled W3/13 or OX-19 (ref. 28, both from Inotech, Switzerland). The TDL of PVG rats contains ~50% B cells (unpublished data). T-depleted TDL (5×10^8) were loaded on to discontinuous Percoll gradients and spun at 1,000g for 30 min. Cells of high buoyant density were collected from the 1.085:1.08 interface and cells of low buoyant density from the 1.065:1.055 interface. Cells (5×10^7) were transferred into carrier (MSH) primed k-allotype distinct recipients that had received 5.5 Gy γ -irradiation one day earlier. Means and standard deviations of five animals per group are presented.

similar antigens have been shown to persist in native form for at least one year^{4,15}. It is known that pathogenic and attenuated viruses can persist for many years in other sites^{16,17}. The results presented here show that memory B cells require periodic contact with antigen for their continued survival in the recirculating pool.

This is not to say that memory B-cell clones contain only activated cells because both large, 'activated' and small, 'resting' cells will transfer a secondary response to carrier-primed adoptive hosts (Fig. 3*b*). The present findings indicate that memory B cells in this resting state are deleted from the recirculating pool within a matter of weeks⁹ unless they meet antigen again. This observation may explain the diverse phenotypic characteristics previously ascribed to memory B cells¹⁸⁻²⁰. Recently Lui *et al.*²¹ showed that hapten-specific cells accumulate in the splenic marginal zone following immunization with DNP-MSH. The authors state that this is a sessile memory pool, but it seems likely that these cells are derived from recirculating cells²² and are also dependent on continued stimulation in follicles.

The surprising loss of the primary host and the donor memory response in 12-week chimaeras (Fig. 1) has at least two possible explanations, one of which was directly testable in our system: either T-cell help was no longer available in 12-week chimaeras or the processing and presentation of soluble antigens to T-helper cells is carried out by activated B cells. The second possibility was tested by transferring 10^7 B cells from DNP-MSH-primed donors into chimaeras 12 weeks after their first cell transfer, together with soluble DNP-MSH. Figure 4 shows that transfer of a small number of specific B cells (donor allotype) brings about the return of the host response. This effect was not mediated as a result of the antibody made by transferred cells facilitating the uptake of the soluble antigen (as immune complex) by other 'professional' presenting cells. This was ruled

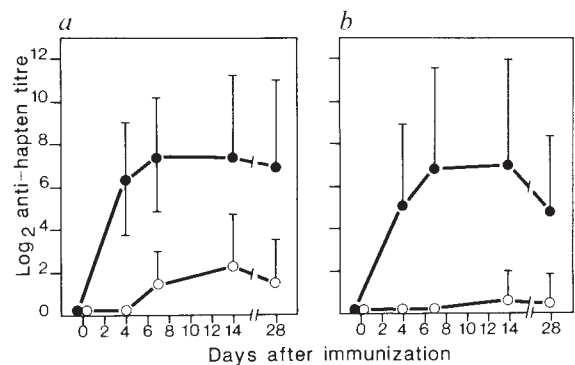


Fig. 4 Activated B cells are required for the processing/presentation of soluble DNP-MSH to T cells *in vivo*. *a*, The host response to soluble DNP-MSH in 12-week chimaeras is restored by transferring 10^7 B cells from DNP-MSH-primed donors at the same time as antigen. *b*, The response in *a* is compared with that elicited by soluble OX-MSH (OX = 2-phenyloxazolone). Means and standard deviation of responses from seven animals are shown.

Methods. Chimaeras were constructed as described in Fig. 1 and left for 12 weeks. At this time they were given with 50 μ g soluble DNP-MSH and OX-MSH, 10^7 TDL depleted of T cells as described in Fig. 3.

out by immunizing 12-week chimaeras with soluble DNP-MSH plus antibodies against DNP-MSH; this protocol elicited no host response. These and other²³ results show that B cells are important in antigen processing and presentation *in vivo*.

What is not clear from these experiments is whether the T-cell help available in 12-week chimaeras is primary or memory. We are at present investigating the persistence of T-cell memory in the absence of antigen using antigen-free transfer of carrier-primed T cells followed by transfer at different times of limiting numbers of hapten-carrier primed B cells together with antigen.

In conclusion, B-cell memory does not reside with very long-lived resting cells, but rather with clones that are maintained over long periods only by continued stimulation with small amounts of antigen persisting on follicular dendritic cells in lymphoid tissues or possibly elsewhere. Unless antigen is sequestered in this way, the increased frequency of antigen-specific precursors that results from immunization is transient. It remains to be seen if all immunological memory can be viewed as an ongoing immune response. In the light of these results it is important to establish the length of time that non-viable antigens remain in lymphoid tissues (as well as the frequency of re-exposure) because this has obvious implications in designing repeat immunization schedules for killed, peptide and recombinant vaccines.

The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche & Co. Ltd, Basel, Switzerland. I am grateful to Drs Harald von Boehmer, David Nemazee, Polly Matzinger and Chris Coleclough for critical reading of the manuscript and Shirley Kirschbaum for typing it.

Received 27 May; accepted 30 September 1988.

1. Strober, S. *Transplant Rev.* **24**, 84-112 (1975).
2. Hood, L. E., Weissman, I. L., Woods, W. B. & Wilson, J. H. in *Immunology*, 2nd edn, 11 (Benjamin/Cummings, Menlo Park, California, 1984).
3. Gray, D., MacLennan, I. C. M. & Lane, P. J. L. *Eur. J. Immunol.* **16**, 641-648 (1986).
4. Tew, J. G., Phipps, R. P. & Mandel, T. E. *Immunol. Rev.* **53**, 175-201 (1980).
5. MacLennan, I. C. M. & Gray, D. *Immunol. Rev.* **91**, 63-85 (1986).
6. Askonas, B. A. & Williamson, A. R. *Eur. J. Immunol.* **2**, 487-493 (1972).
7. Lane, P. J. L., Gray, D., Oldfield, S. & MacLennan, I. C. M. *Eur. J. Immunol.* **16**, 1569-1575 (1986).
8. Gray, D. *Immunology* **65**, 73-79 (1988).
9. Gray, D. *J. exp. Med.* **167**, 805-816 (1988).
10. Askonas, B. A., Cunningham, A. J., Kreth, H. W., Roelants, G. E. & Williamson, A. R. *Eur. J. Immunol.* **2**, 494-498 (1972).
11. Feldbush, T. L. *Cell Immunol.* **8**, 435-444 (1973).
12. Celada, F. J. *J. exp. Med.* **125**, 199-211 (1967).
13. Mandel, T. E., Phipps, R. P., Abbot, A. & Tew, J. G. *Immunol. Rev.* **53**, 29-57 (1980).

14. Tew, J. G. & Mandel, T. E. *J. Immun.* **120**, 1063-1069 (1978).
15. Tew, J. G. & Mandel, T. E. *Immunology* **37**, 69-76 (1979).
16. Mims, C. A. *Pathogenesis of Infectious Disease* 3rd edn, 132 (Academic, London, 1987).
17. Frankel, L. D. & Bellanti, J. A. in *The Immunology of Human Infection Part II* (eds Nahamias, A. J. & O'Reilly, R. J.) 135 (Plenum, London, New York, 1982).
18. Black, S. J., Tokuhisa, T., Herzenberg, L. A. & Herzenberg, L. A. *Eur. J. Immun.* **10**, 846-851 (1980).
19. Zan-Bar, I., Strober, S. & Vitetta, E. S. *J. Immun.* **123**, 925-930 (1979).
20. Hayakawa, K., Ishii, R., Yamasaki, K., Kishimoto, T. & Hardy, R. R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1379-1383 (1987).
21. Lui, Y.-J., Oldfield, S. & MacLennan, I. C. M. *Eur. J. Immun.* **18**, 355-362 (1988).
22. Kumararatne, D. S. & MacLennan, I. C. M. *Eur. J. Immun.* **11**, 865-869 (1981).
23. Ron, Y. & Sprent, J. *J. Immun.* **138**, 2848-2856 (1987).
24. Hunt, S., V. & Fowler, M. H. *Cell Tissue Kinet.* **14**, 445-464 (1981).
25. Gray, D., Chassoux, D., MacLennan, I. C. M. & Bazin, H. *Clin. exp. Immun.* **60**, 78-86 (1985).
26. Mäkelä, O., Kaartinen, M., Pelkonen, J. L. T. & Karjalainen, K. *J. exp. Med.* **148**, 1644-1660 (1978).
27. Williams, A. F., Galfré, G. & Milstein, C. *Cell* **12**, 663-637 (1977).
28. Dallman, M. J., Mason, D. W. & Webb, M. *Eur. J. Immun.* **12**, 511-518 (1982).

Positive and negative selection of an antigen receptor on T cells in transgenic mice

William C. Sha, Christopher A. Nelson,
Rodney D. Newberry, David M. Kranz*,
John H. Russell† & Dennis Y. Loh

Howard Hughes Medical Institute,
Washington University School of Medicine,
Departments of Medicine, Genetics, Microbiology and
Immunology, and † Pharmacology, St Louis, Missouri 63110, USA
* Department of Biochemistry, University of Illinois, Urbana,
Illinois 61801, USA

The T-cell repertoire found in the periphery is thought to be shaped by two developmental events in the thymus that involve the antigen receptors of T lymphocytes. First, interactions between T cells and major histocompatibility complex (MHC) molecules select a T-cell repertoire skewed towards recognition of antigens in the context of self-MHC molecules¹⁻⁵. In addition, T cells that react strongly to self-MHC molecules are eliminated by a process called self-tolerance⁶⁻¹⁰. We have recently described transgenic mice expressing the $\alpha\beta$ T-cell receptor from the cytotoxic T lymphocyte 2C (ref. 11). The clone 2C was derived from a BALB.B (H-2^b) anti-BALB/c (H-2^d) mixed lymphocyte culture and is specific for the L^d class I MHC antigen. In transgenic H-2^b mice, a large fraction of T cells in the periphery expressed the 2C T-cell receptor. These T cells were predominantly CD4⁺CD8⁺ and were able to specifically lyse target cells bearing L^d. We now report that in the periphery of transgenic mice expressing L^d, functional T cells bearing the 2C T-cell receptor were deleted. This elimination of autoreactive T cells appears to take place at or before the CD4⁺CD8⁺ stage in thymocyte development. In addition, we report that in H-2^s mice, a non-autoreactive target haplotype, large numbers of CD8⁺ T cells bearing the 2C T-cell receptor were not found, providing strong evidence for the positive selection of the 2C T-cell receptor specificity by H-2^b molecules.

Male founder, P, an H-2^b animal derived from injection of eggs obtained from mating (C57L × SJL)F₁ H-2^{b/s} mice, gave rise to transgenic offspring containing either eight copies of the β and one copy of the α transgene, denoted as type B, or those containing two copies of the β and four copies of the α transgene, denoted as type A. The 2C T-cell receptor (TCR) can be detected by two monoclonal antibodies, F23.1 and 1B2. F23.1 recognizes the β chain of the 2C TCR and two other members of the V β 8 gene family^{12,13}. 1B2 is a clonotypic antibody specific to the 2C TCR (ref. 14). Activated 1B2⁺ T cells from transgenic H-2^b mice were able to lyse lymphoblast targets from BALB/c (H-2^d), but not from SJL (H-2^s) or dm2 (a BALB/c L^d loss mutant)¹⁵. So in order to study the fate of autoreactive T cells, type B H-2^{b/d} animals were generated by crossing founder P to

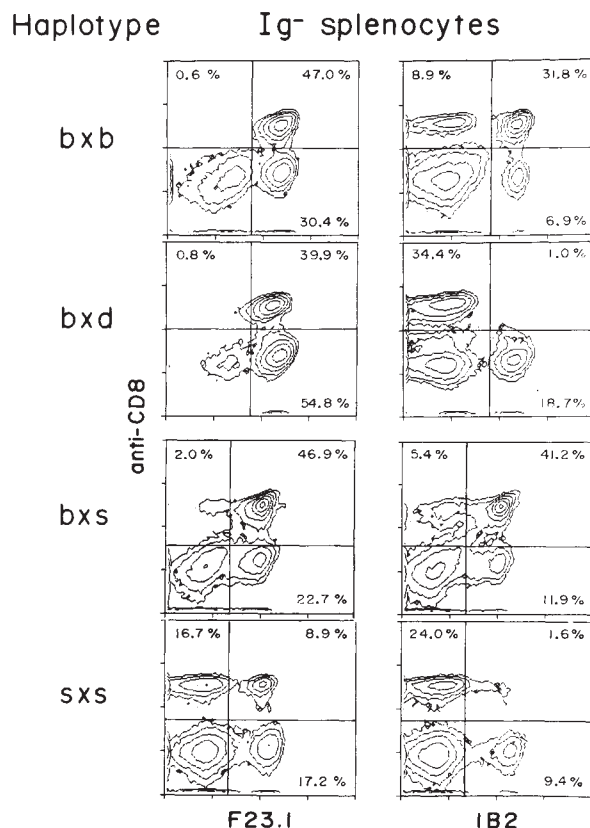


Fig. 1 Selective depletion of CD8⁺1B2⁺ immunoglobulin-negative spleen cells in transgenic H-2^{b/d} and H-2^s animals as shown by two-colour flow cytometric analysis of F23.1 and 1B2 expression on CD8⁺ T cells. The H-2^b and H-2^{b/d} mice were type B littermates. The H-2^s and H-2^s mice were type A littermates. Haplotypes were determined by Southern blot analysis using a 0.64-kilobase BamHI fragment 5' to the H-2K^b gene¹⁷. An L^d-specific antibody, 30-5-7, was also used to screen H-2^s from H-2^{b/d} mice¹⁸. As we described elsewhere¹¹, very few 1B2⁺ T cells were CD4⁺ in H-2^b animals. CD4⁺1B2⁺ cells were absent in H-2^{b/d} mice and were <4% of immunoglobulin-negative splenocytes in H-2^s mice. In the H-2^b mouse, 19.2% of immunoglobulin-negative spleen cells were CD4⁺F23.1⁺, 1.1% were CD4⁺F23.1⁻, and 0.8% were CD4⁺1B2⁺. In the H-2^{b/d} mouse, 36.2% were CD4⁺F23.1⁺, 0.9% were CD4⁺F23.1⁻, and 0.5% were CD4⁺1B2⁺. In the H-2^s mouse, 20.3% were CD4⁺F23.1⁺, 9.6% were CD4⁺F23.1⁻, and 7.7% were CD4⁺1B2⁺. In the H-2^s mouse, 8.7% were CD4⁺F23.1⁺, 31.4% were CD4⁺F23.1⁻, and 1.6% were CD4⁺1B2⁺. Similar flow-cytometry data of peripheral transgene expression was obtained from five other sets of type B, H-2^b and H-2^{b/d} littermates. Six other type A H-2^s mice, derived from two other male H-2^{b/s} offspring of founder P, were also analysed with the same results. Other figures and tables present data generated from the same H-2^b, H-2^{b/d}, and H-2^s mice shown here.

Methods. Splenocytes were isolated by squeezing whole spleen through a wire screen. Red blood cells were removed by ammonium chloride lysis, and passed over a Sci Can T-cell-recovery column (Sci Can Diagnostics, Edmonton) to remove immunoglobulin-positive cells. Flow cytometry analysis was performed as described¹¹. First-step reagents used were biotinylated F23.1, biotinylated 1B2, and 53-6.72 (rat anti-mouse CD8) (ref. 19). Second-step reagents used were streptavidin-phycoerythrin and fluorescein isothiocyanate conjugated goat anti-rat immunoglobulin cross-absorbed against mouse immunoglobulin (Southern Biotechnology, Alabama). 70,000 to 100,000 fluorescence-activated cell sorting events were analysed for each contour plot using a Becton-Dickinson FACS 440. In these plots, streptavidin-phycoerythrin, and fluorescein isothiocyanate fluorescence are plotted on the x and y axes respectively. Contour intervals used were 10, 30, 90, 270 and 810 cells per pixel. The x and y coordinates that defined negative cells were selected as the intersection of positive and control negative profiles in each parameter.

Table 1 Functional transgene is not expressed in H-2^d animals

Responder	Cells recovered $\times 10^{-7}$	E:T	Specific lysis (%)	
			BALB/c	dm2
H-2 ^b	1.5	4:1	58 $\pm$ 2	6 $\pm$ 1
		20:1	59 $\pm$ 1	4 $\pm$ 2
H-2 ^{b/d}	0.2	17:1	13 $\pm$ 0	13 $\pm$ 1

Splenocytes from transgenic littermates of the indicated H-2 haplotypes were stimulated for 5 days in the presence of irradiated BALB/c (H-2^d) splenocytes and IL-2-containing conditioned medium. The latter was added to bypass the necessity for growth or differentiation factors potentially not made by the 1B2⁺CD8⁺ cells. The cells were collected and tested for lytic activity against lymphoblast targets from BALB/c or its H-2L^d-deletion mutant, dm2. The responding populations were the same as depicted at culture initiation in Fig. 1. The weak activity against BALB/c is not L^d-specific as there is comparable activity against dm2. This represents one of four similar experiments. Responding splenocytes (10^7) were mixed with 2×10^7 irradiated (2,000 rad) BALB/c splenocytes in 10 ml RPMI-1640 medium supplemented with supernatant (10 U ml^{-1} final IL-2 concentration) from concanavalin A-stimulated rat splenocytes and fetal calf serum. α -Methyl mannoside was added to a final concentration of 50 mM to block the activity of residual concanavalin A. After 5 days the cells were collected, cleared of debris by Ficoll-Hypaque centrifugation and assayed for 3 h against 10^4 concanavalin A lymphoblasts which had been cleared of debris as above and labelled with ^{51}Cr in the presence of 50 mM α -methyl mannoside. The lytic assay medium was 200 μl of the stimulating culture medium minus the concanavalin A supernatant and α -methyl mannoside.

(BALB/c $\times$ C57B1/6)F₁ H-2^{b/d} mice. Unlike transgenic H-2^b littermates, H-2^{b/d} animals did not have functional T cells bearing the 2C TCR. Flow-cytometric analysis showed that it was predominantly CD4⁺CD8⁺ cells that expressed the 2C TCR. In type B H-2^b animals, almost all peripheral CD4⁺ or CD8⁺ T cells were F23.1⁺, indicating that essentially every T cell expressed the 2C β chain. A smaller percentage of T cells (55 to 80%) were also 1B2⁺. Only very few of the T cells expressing the 2C TCR were CD4⁺, being mostly CD4⁺CD8⁺. In type B H-2^{b/d} mice, again almost all peripheral CD4⁺ or CD8⁺ T cells were F23.1⁺. Expression of the 2C TCR, however, was limited to CD4⁺CD8⁺ cells, indicating that T cells with CD4 or CD8 accessory molecules expressed only the β chain of the 2C TCR, presumably with a different α chain. Thus, compared with transgenic H-2^b mice, 1B2⁺CD8⁺ T cells were selectively depleted in the periphery.

To test whether T cells expressing the 2C TCR would arise in comparably high numbers in a non-autoreactive H-2 environment apart from H-2^b, transgenic H-2^s animals were generated by backcrossing founder P twice to (C57L $\times$ SJL)F₁ mice. In H-2^s animals, most T cells did not express the β transgene on the cell surface as only 25–30% of T cells were F23.1⁺. Interestingly, the peripheral expression of the 2C TCR in H-2^s animals was similar to H-2^{b/d} animals. 2C TCR-bearing cells in the periphery were predominantly CD4⁺CD8⁺. In contrast, the 2C TCR was evident on CD8⁺ T cells in five H-2^{b/s} mice. Expression of F23.1 and 1B2, as well as the skewing of clonotype to CD8⁺ cells, however, was more variable than it was in H-2^b mice. The 2C TCR represents a potential allo-specific clonotype for both H-2^b and H-2^s, but is only expanded in the H-2^b environment. Because founder P and offspring contained genetic components from both C57L and SJL strains, it is not possible to state categorically that loci outside the MHC did not influence the peripheral expression of CD8⁺ T cells bearing the 2C TCR. Nevertheless, we have observed no exceptions to the peripheral expression (shown in Fig. 1) in seven H-2^s and more than twelve H-2^b mice analysed. This observation provides evidence for positive selection of the 2C TCR by H-2^b elements.

Thus, in contrast to H-2^b and H-2^{b/s} mice, peripheral expression of the 2C TCR in H-2^{b/d} and H-2^s mice was predominantly on CD4⁺CD8⁺ cells, which were Thy1⁺ (data not shown). In H-2^{b/d} mice, clonal deletion of 1B2⁺CD8⁺ T cells was the likely explanation, as these cells have been shown to be functionally active against L^d. This was confirmed by stimulating splenocytes from H-2^b and H-2^{b/d} littermates (Fig. 1) with BALB/c spleen cells. The functional deletion of L^d reactive cells in H-2^{b/d} mice (Table 1) was reflected not only in the lack of L^d-specific lytic activity, but also in the difference in cell recoveries from the cultures of spleen cells. In response to H-2^k alloantigens, type B H-2^{b/d} animals responded to the same extent or better than type B H-2^b littermates (data not shown).

Functional 1B2⁺CD8⁺ T cells could be detected in H-2^s mice after stimulation with BALB/c spleen cells (Table 2), but not in an auto-reactive L^d-containing environment. In H-2^{b/s} mice, cytolytic T cells bearing the 2C TCR completely dominated the response to H-2^d alloantigens. In H-2^s mice, however, only a third of cells responding to H-2^d alloantigens were 1B2⁺CD8⁺. The difference in responses of H-2^{b/s} and H-2^s mice was evident in the ability of the bulk cultures to lyse dm2 targets and the ability of 1B2 to inhibit lysis of BALB/c targets. This experiment suggests that the failure to express 1B2⁺CD8⁺ cells in the periphery of the H-2^s mice is not the result of negative selection as it is in the H-2^{b/d} mice. It also supports the idea that the high levels of 1B2⁺CD8⁺ in H-2^b-containing animals are the result of a positive selection of these cells by H-2^b elements. In this transgenic model the presence of low numbers of functional 1B2⁺ T cells in H-2^s mice raises the issue of whether H-2^s elements are selecting weakly for cells bearing the 2C TCR, or whether positive selection is not required for all T-cell differentiation, but rather represents a skewing mechanism for the amplification of particular clonotypes.

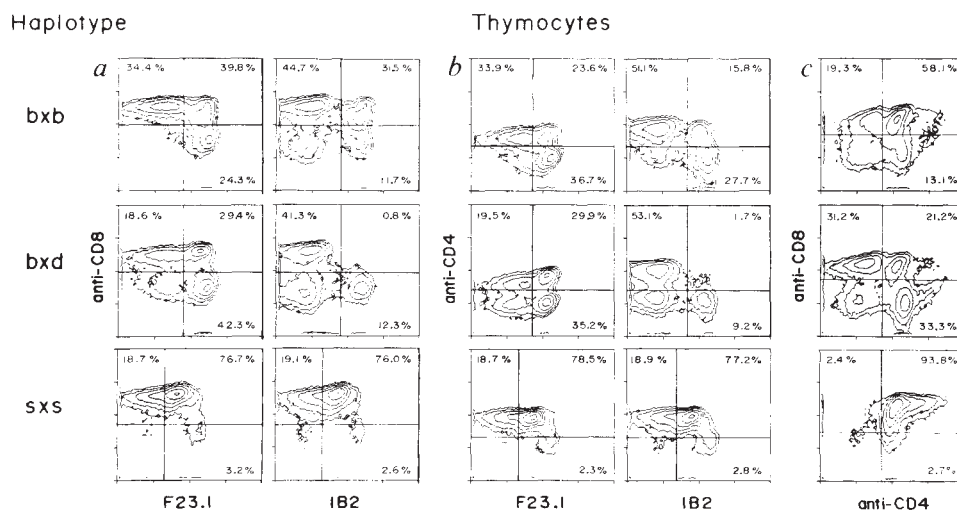
In H-2^b animals, the majority of thymocytes expressed the 2C TCR. Skewing towards CD8⁺ T cells in the periphery was evident in an increased population of CD4⁺CD8⁺ thymocytes and a decreased population of CD4⁺CD8⁺ thymocytes relative to nontransgenic control mice. In type B H-2^{b/d} animals, the 2C TCR was expressed predominantly on CD4⁺CD8⁺ thymocytes. Depletion of CD4⁺CD8⁺ thymocytes was clonotype-specific, as judged by the presence of a significant population of CD4⁺CD8⁺ thymocytes and the presence of populations of F23.1⁺ CD8⁺ and CD4⁺ thymocytes. This strongly suggests that autoreactive T cells expressing the 2C TCR were selectively eliminated at or before the double-positive stage in thymocyte development.

In H-2^s animals, the small percentage of functional 2C TCR-bearing T cells in the periphery contrasted with the large percentage of thymocytes (>70%) expressing the 2C TCR. Thymuses of H-2^s animals contained $2\text{--}4 \times 10^8$ thymocytes, which was several times the number isolated from either H-2^b or H-2^{b/d} mice. Most of these thymocytes were of the nonmature CD4⁺CD8⁺ phenotype. The small percentage of 1B2⁺CD8⁺ T cells in the periphery, and the relatively small populations of CD4⁺CD8⁺ and CD4⁺CD8⁺ thymocytes in H-2^s mice, indicate that immature 2C TCR-bearing thymocytes matured at a lower frequency in H-2^s than in H-2^b mice. Thus, although 1B2⁺ T cells in the periphery of H-2^s and H-2^{b/d} mice were predominantly CD4⁺CD8⁺, 1B2⁺ thymocytes in H-2^s mice were largely CD4⁺CD8⁺, whereas 1B2⁺ thymocytes in H-2^{b/d} mice were CD4⁺CD8⁺.

Our results indicate that the H-2 environment in which T cells differentiate can alter the T-cell repertoire by both selective amplification and depletion of particular T-cell specificities. In agreement with other reported results^{6–10} our data demonstrate that clonal deletion of autoreactive T cells is an important mechanism of self-tolerance. As also observed by Kisielow *et al.*¹⁰, T cells that expressed the 2C TCR and which were CD4⁺CD8⁺ apparently survived selection and were functionally inactive in the periphery. In the thymuses of animals in which

Fig. 2 Two-colour flow cytometric analysis of F23.1, 1B2, CD4, and CD8 expression on the thymocytes of H-2^b, H-2^{b/d} and H-2^s mice shown in Fig. 1. Similar results were obtained from five other sets of H-2^b and H-2^{b/d} littermates and from six other H-2^s mice.

Methods. Thymocytes were isolated by squeezing the thymus through a wire screen. Red blood cells were removed by ammonium chloride lysis. Flow cytometry was performed as described in Fig. 1. Additional first-step reagents used were RL172.4 (rat anti-mouse CD4) (ref. 20) for *b* and AD4(15) (mouse anti-mouse CD8; Cedarlane, NY) and GK1.5 (rat anti-mouse CD4) (ref. 21) for *c*. Additional second-step reagents used were cross-absorbed fluorescein isothiocyanate-conjugated goat anti-rat IgG and streptavidin-phycoerythrin conjugated goat anti-mouse IgM (Southern Biotechnology) for *c*.



2C TCR-bearing cells would be autoreactive, very few CD4⁺ or CD8⁺ clonotype positive cells were found. Elimination of CD4⁺CD8⁺ cells appeared to be specific to autoreactive T cells, as thymuses of these animals had a sizeable population of 1B2⁺CD4⁺CD8⁺ thymocytes.

Our results on negative selection in transgenic animals differ from those of Kisielow *et al.*¹⁰ in two important ways. First, thymuses from type B H-2^{b/d} animals did not differ substantially in size or number of thymocytes from transgenic H-2^b littermates. Second, CD4⁺CD8⁺ thymocytes from these mice expressed the same surface densities of CD4 and CD8 accessory molecules as CD4⁺CD8⁺ thymocytes from transgenic H-2^b mice. It is unlikely that these differences result from the particular receptor chosen for study, because both the 2C and B6.2.16 clones lyse targets in a CD8-dependent fashion. Another potential explanation for these differences is the variation in ability of transgenic animals to produce clonotype-negative T cells. Type A H-2^b mice generated from founder P consistently expressed the 2C TCR on a large percentage of total peripheral T cells. In two type A H-2^b littermates of the H-2^{b/d} animals tested, 90 and 98% of F23.1⁺ cells were 1B2⁺. In contrast, in type B H-2^b animals, a more variable percentage, 70±12% (*n*=4), of F23.1⁺ T cells were 1B2⁺. To investigate whether levels of clonotype expression could affect clonal deletion of

autoreactive T cells, we analysed three type A H-2^{b/d} mice. Spleen cells from these mice did not respond to L^d-bearing cells in assays of functional activity, and they also did not respond to third party targets as well as transgenic control littermates (data not shown). These mice had small thymuses, with reduced numbers of thymocytes compared with type A H-2^b littermates, and had clonotype-positive cells in both the thymus and spleen-bearing lower levels of CD4 and CD8. Thus, our results from type A H-2^{b/d} mice are similar to those reported for male transgenic mice by Kisielow *et al.*¹⁰. Data from the type B H-2^{b/d} mice indicate that a more complete elimination of clonotype-positive T cells bearing accessory molecules can occur in transgenic mice where clonotype expression is not as dominant.

The results from our experiments testing for positive selection are consistent with the model suggested by Bevan and Hunig¹⁶, in which the T-cell repertoire is skewed by selective amplification of thymocytes capable of interacting with self-MHC antigens. There are several reasons why selection against an autoreactive T-cell specificity is an unlikely explanation for the small numbers of 1B2⁺CD8⁺ cells found in the periphery of H-2^s mice. First, T cells isolated from an H-2^b animal were able to lyse BALB/c but not SJL concanavalin A lymphoblast targets. Second, the large population of 1B2⁺CD4⁺CD8⁺ thymocytes in H-2^s thymuses is inconsistent with the depletion of double-positive thymocytes bearing autoreactive receptors demonstrated here and by Kisielow *et al.*¹⁰. Third, H-2^{b/s} mice contained a population of 1B2⁺CD8⁺ T cells that was clearly detectable by flow cytometry and which completely dominated responses to H-2^d alloantigens. Unlike transgenic H-2^{b/d} mice, transgenic H-2^{b/s} mice contained functional 1B2⁺CD8⁺ T cells. This indicates that the lack of 1B2⁺CD8⁺ T cells in the periphery of H-2^s mice results from the absence of H-2^b elements, rather than from the presence of H-2^s elements. Thus, the alloreactive specificity of the 2C clone is apparently expanded in an H-2^b, but not in an H-2^s, thymus because it can interact successfully with H-2^b, but not H-2^s, MHC antigen(s). This interaction appears to require both $\alpha\beta$ heterodimer and CD4/CD8 accessory molecules, as judged by the selective expression of the 2C TCR on CD8⁺ T cells in H-2^b mice¹¹.

We thank D. McCulley for technical assistance, T. Hansen for dm2 mice and 30-5-7 antibody, G. Elliott and B. Smith for performing flow cytometry, E. Weiss for the H-2K^b probe and P. Allen for helpful comments on the manuscript. W.S. is supported by a Medical Scientist Training grant, J.H.R. by the NIH and D.Y.L. by the Howard Hughes Medical Institute.

Table 2 Functional transgene bearing cells can be expanded in H-2^s animals

Responder	Cells recovered × 10 ⁻⁷	E:T	Specific lysis (%)			
			BALB/c	1B2 plus BALB/c	dm2	1B2 plus dm2
H-2 ^{b/s}	1.9	0.4	18±0	0±1	0±1	1±1
		2	47±2	1±0	1±1	2±0
		10	59±1	4±0	0±0	5±2
H-2 ^s	1.0	0.4	13±1	4±0	3±0	2±1
		2	40±1	14±1	11±0	12±0
		10	62±2	39±1	31±2	29±0

Splenocytes from transgenic littermates of the indicated H-2 haplotypes were stimulated with BALB/c splenocytes as described in Table 1 and assayed against BALB/c and dm2 lymphoblast targets either alone or in the presence of the anti-clonotype 1B2 (10 µg ml⁻¹). Cells from both cultures were also stained and examined for 1B2⁺CD8⁺ cells. The result was that 95% and 31% of the cells from the H-2^{b/s} and H-2^s cultures respectively were 1B2⁺CD8⁺. These were the same H-2^s cells that were <2% 1B2⁺CD8⁺ depicted in Fig. 1 at culture initiation. This represents one of three similar experiments.

Received 25 July; accepted 23 September 1988.

1. Born, W. *et al. J. Immun.* **138**, 999-1008 (1987).
2. McDuffie, M., Born, W., Marrack, P. & Kappler, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8728-8733 (1986).
3. Kruisbeek, A. *et al. J. exp. Med.* **161**, 1029-1047 (1985).
4. Bevan, M. *Nature* **269**, 417-419 (1977).
5. Zinkernagel, R. *et al. J. exp. Med.* **161**, 1029-1047 (1985).
6. Kappler, J. *et al. Cell* **49**, 263-271 (1987).
7. Kappler, J., Roehm, N. & Marrack, P. *Cell* **49**, 273-280 (1987).
8. Kappler, J., Staerz, U., White, J. & Marrack, P. *Nature* **332**, 35-40 (1988).
9. MacDonald, H. *et al. Nature* **332**, 40-45 (1988).
10. Kieselow, P., Bluthmann, H., Staerz, U., Steinmetz, M. & von Boehmer, H. *Nature* **333**, 742-746 (1988).
11. Sha, W. C. *et al. Nature* **335**, 271-274 (1988).
12. Staerz, U., Rammensee, H., Benedetto, J. & Bevan, M. *J. Immun.* **134**, 3994-4000 (1985).
13. Behlke, M. *et al. J. exp. Med.* **165**, 257-262 (1987).
14. Kranz, D., Sherman, D., Sitkovsky, M., Pasternack, M. & Eisen, H. *Proc. natn. Acad. Sci. U.S.A.* **81**, 573-577 (1984).
15. Rubocki, R., Hansen, T. & Lee, D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9606-9610 (1986).
16. Bevan, M. & Hunig, T. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1843-1847 (1981).
17. Weiss, E. *et al. Nature* **310**, 650-655 (1984).
18. Ozato, K., Hansen, T. & Sachs, D. *J. Immun.* **125**, 2473-2477 (1980).
19. Ledbetter, J. & Herzenberg, L. *Immun. Rev.* **47**, 63-90 (1979).
20. Ceredig, R., Lowenthal, J., Nabholz, M. & MacDonald, H. *Nature* **314**, 98-100 (1985).
21. Dyalinas, D. *et al. Immun. Rev.* **74**, 28-56 (1983).

Engagement of the CD4 molecule influences cell surface expression of the T-cell receptor on thymocytes

Susan A. McCarthy, Ada M. Kruisbeek*, Ingeborg K. Uppenkamp, Susan O. Sharrow & Alfred Singer

Experimental Immunology Branch and
*Biological Response Modifiers Programme,
National Cancer Institute, National Institutes of Health,
Bethesda, Maryland 20892, USA

The intrathymic differentiation process by which precursor cells derived from the bone marrow develop into immuno-competent T lymphocytes is poorly understood. Most thymocytes express both CD4 and CD8 accessory molecules¹, yet little is known about either the function of these molecules or the responsiveness of the CD4⁺8⁺ double positive thymocytes that bear them²⁻⁷. Here, we address the possibility that CD4 engagement influences T-cell receptor (TCR) expression on developing thymocytes. We engaged CD4 molecules on murine thymocytes by *in vivo* injection of an anti-CD4 monoclonal antibody, which reduced the surface expression of CD4 on CD4⁺ thymocytes. More importantly, CD4 engagement also affected TCR expression on CD4⁺ thymocytes, but the effect on CD4⁺8⁺ double positive and CD4⁺8⁻ single positive thymocytes was very different. CD4⁺8⁺ thymocytes responded to CD4 engagement by dramatically increasing surface expression of TCR, whereas CD4⁺8⁻ thymocytes decreased surface expression of TCR. These results demonstrate that the effect of CD4 engagement on TCR expression is dependent upon the developmental state of the responding thymocyte, and, most interestingly, results in increased TCR expression by double positive thymocytes.

We initially investigated the responses of murine thymocytes to anti-CD4 monoclonal antibody (mAb) by injecting the mAb into newborn C57BL/6 (B6) mice for seven days and assessing CD3/TCR expression on double positive thymocytes by immunofluorescence and flow cytometry. These thymocytes exhibited dramatic increases in cell surface CD3 (Fig. 1c) and TCR (Fig. 1d; V β 8, expressed on about 15% of CD3⁺ cells) in response to anti-CD4 mAb treatment. With regard to its effects on other thymocyte surface proteins, anti-CD4 mAb substantially reduced surface expression of its ligand, CD4 (Fig. 2a, i),

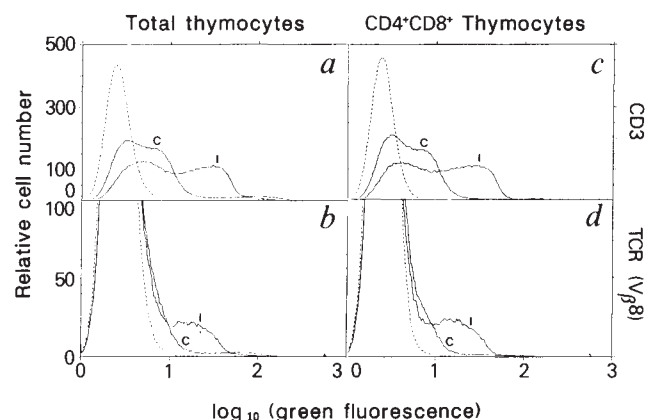


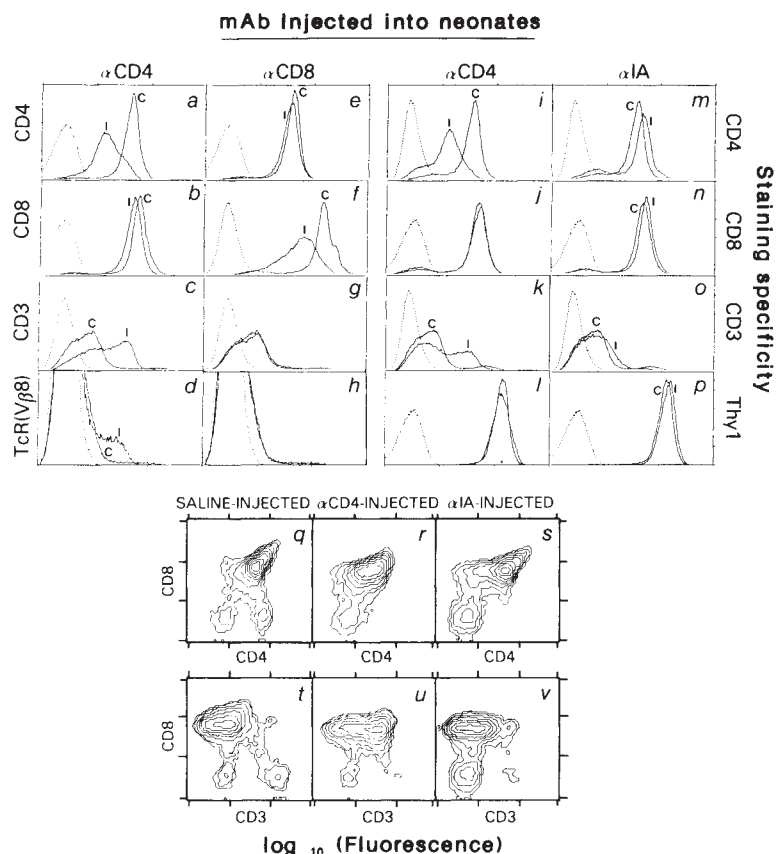
Fig. 1 Double positive thymocytes respond to *in vivo* anti-CD4 mAb treatment by increasing their cell-surface expression of the CD3/TCR complex. Thymocytes from mice injected with anti-CD4 mAb or saline were stained and analysed by three-colour flow cytometry. Single-colour staining histograms of CD3 (a and c) and TCR (V β 8; b and d) are shown for double positive thymocytes (c and d) and total thymocytes (a and b). Each single colour histogram shows staining of thymocyte cell populations from either control saline-injected mice (solid line, designated 'C') or from experimental anti-CD4 mAb-injected mice (solid line, designated 'I'), by either a negative control antibody Leu4 (dotted line) or by antibodies with the indicated specificities. The median relative staining intensities of cells from anti-CD4 mAb-injected mice compared to cells from control mice (defined as 100% for each panel) were: a, 327%; c, 379%. These increases were not merely due to an increase in overall cell size, as the forward light scatter of CD4⁺8⁺ thymocytes from control and injected animals were virtually identical. In addition, these increases were also not due simply to selective loss of CD4⁺8⁺ thymocytes with low level TCR expression because cell recovery of thymocytes were similar from control and anti-CD4 injected mice.

Methods. B6 pregnant female mice were purchased from the Jackson Laboratory, Bar Harbor, Maine. Starting on day one after birth, neonates were injected daily with 100 μ g purified GK1.5 anti-CD4 mAb¹⁹ or saline, intraperitoneally. On day 7, thymocytes were prepared and stained in sequence for CD3 or TCR, CD4, and CD8, and analysed by three-colour flow cytometry. CD3 staining in green was with 145-2C11 mAb²⁰ conjugated with fluorescein isothiocyanate (FITC), TCR staining in green was with F23.1-FITC (ref. 21), CD4 staining in orange was with GK1.5 culture supernatant plus MAR18.5 (a mouse anti-rat κ -chain mAb²²) conjugated with phycoerythrin (PE), and CD8 staining in red was with 53-6.72-Biotin (anti-Lyt2, Becton Dickinson Immunocytometry Systems (BDIS), Mountain View, California) plus Texas Red streptavidin (Bethesda Research Laboratories, Maryland). Negative control stains were Leu4-FITC, Leu15-PE and Leu1-Biotin, which are specific for human cells, and do not stain murine T cells. Viable cells (250,000), as defined by forward light scatter intensity, were analysed for each sample on a modified FACS II (BDIS) interfaced to a PDP 11/84 computer (Digital Equipment Corp., Maynard, Maine), as previously described¹⁴. Data were collected in list mode using logarithmic amplification provided by a three-decade logarithmic amplifier constructed using an NIH-modified design of R. Hiebert (Los Alamos Scientific Laboratories, Los Alamos, New Mexico). Identification of CD4⁺8⁺ thymocytes was based on CD4 staining in orange and CD8 staining in red, and data were software-gated to generate the green staining histograms (Leu4, CD3 and TCR) of those cells. For quantitative comparisons of staining intensity, median channel values of fluorescence intensity were converted to linear units using empirically-derived standard calibration curves constructed for each amplifier used.

without significantly affecting either CD8 (Fig. 2b, j) or Thy1 (Fig. 2l). In contrast, anti-CD8 mAb did not increase CD3/TCR expression on CD4⁺8⁺ thymocytes (Fig. 2g, h), although it did reduce surface expression of its ligand, CD8 (Fig. 2f). The reasons for the different effects on TCR expression by double

Fig. 2 Effects of *in vivo* anti-CD4, anti-CD8, and anti-Ia mAb treatments on thymocytes. Murine thymocytes were analysed in two separate experiments, experiment 1 (a-h) and experiment 2 (i-p). Thymocytes from mice injected with anti-CD4 mAb (panels a-d, i-l, r and u), anti-CD8 mAb (e-h), anti-Ia mAb (m-p, s and v) or saline (a-p, q and t) were stained and analysed by flow cytometry. Single-colour histograms from two-colour analyses are shown for CD4 (a, e, i and m), CD8 (b, f, j and n), CD3 (c, g, k and o), TCR ($V_{\beta}8$; d and h) and Thy1 (l and p). Each single-colour histogram shows staining of thymocyte cell populations from either control saline-injected mice (solid line, designated 'C') or from experimental mAb-injected mice (solid line, designated 'I'), by either a negative control antibody (dotted line) or by antibodies with the indicated specificities. Two-colour contour analyses are shown for CD8 versus CD4 staining (q-s) and CD8 versus CD3 staining (t-v) of thymuses from saline-injected (q and t), anti-CD4 mAb-injected (r and u) and anti-Ia mAb-injected (s and v) mice.

Methods. B6 pregnant female mice were purchased from the Jackson Laboratory. Starting on day 1 after birth, neonates were injected intraperitoneally daily with 100 μ g (days 1-7) or 200 μ g (days 8-9) purified GK1.5 anti-CD4 mAb, purified 2.43 anti-CD8 mAb²³, purified Y-3P anti-Ia^b mAb²⁴, or saline. On day 10, thymocytes were prepared and stained for CD4, CD8, CD3, TCR, and Thy1, and then analysed by two-colour flow cytometry. CD4 staining was with GK1.5 culture supernatant followed by MAR18.5-Biotin plus Texas Red streptavidin (a) or MAR18.5-FITC (i, m and q-s), or with H129.19-FITC anti-CD4 mAb²⁵ (e). CD8 staining was with 53-6.72-Biotin plus Texas Red streptavidin (b, j, n and q-v), or with purified 2.43 mAb followed by MAR18.5-FITC (f). CD3 staining was with 145-2C11-FITC, TCR ($V_{\beta}8$) staining was with F23.1-FITC, and Thy1 staining was with 30-H12-Biotin (anti-Thy1.2 mAb, BDIS). Negative control stains were Leu4-FITC and Leu1-Biotin. Fluorescence data were collected using logarithmic amplification on 5×10^4 viable cells, as determined by forward light scatter and propidium iodide exclusion. For quantitative comparisons of staining intensity, median channel values of fluorescence intensity were converted to linear units using empirically derived standard calibration curves.



positive thymocytes of anti-CD4 and anti-CD8 mAbs are not yet clear.

Increased surface expression of TCR on double positive thymocytes must be either a direct response of the double positive cells to anti-CD4 mAb, or a secondary response of double positive cells to the depletion of $CD4^+8^-$ thymocytes by chronic anti-CD4 treatment. This latter possibility arises because mice chronically injected with anti-CD4 mAb are devoid of $CD4^+8^-$ cells, as can be seen in Fig. 2q-u ($CD4^+8^-$ cells are CD3-bright, and are located in the lower right quadrant of Fig. 2t but are missing in Fig. 2u). In that case, any *in vivo* treatment that depletes $CD4^+8^-$ cells would be expected to also increase CD3/TCR expression on $CD4^+8^+$ thymocytes. However, anti-Ia mAb treatment did not increase thymocyte CD3 expression (Fig. 2k, o), even though both anti-Ia and anti-CD4 treatments depleted $CD4^+8^-$ thymocytes (Fig. 2q-v). Furthermore, anti-Ia mAb did not interfere with the inductive effect of anti-CD4 mAb in neonatal mice simultaneously treated with both anti-Ia and anti-CD4 mAbs (data not shown). Thus, these results indicate that anti-CD4 mAb acts directly on double positive thymocytes to increase CD3/TCR expression. This conclusion is further supported by the observation that $CD4^+8^+$ thymocytes respond within one day to anti-CD4 mAb treatment (Table 1, experiment 1), even before anti-CD4 mAb depletes $CD4^+8^-$ thymocytes.

Because double positive thymocytes responded rapidly to anti-CD4 mAb treatment, it was possible to compare the acute effects of anti-CD4 mAb on CD3 expression by $CD4^+8^+$ and $CD4^+8^-$ cells within a single thymocyte population. We analysed

double positive and single positive thymocytes from 10-day-old C3H mice injected for two days with anti-CD4 mAb. The effect of anti-CD4 mAb on double positive cells was to reduce CD4 surface expression (Fig. 3a) and increase CD3 surface expression (Fig. 3b); whereas the effect of anti-CD4 mAb on $CD4^+8^-$ cells from the same thymus preparation was to decrease both CD4 (Fig. 3c) and CD3 expression (Fig. 3d). This dichotomy between increased CD3 expression on double positive thymocytes but decreased CD3 expression on single positive thymocytes, even though cell-surface CD4 is reduced on both cell types, demonstrates that the effect of CD4 engagement on expression of CD3/TCR is determined by the developmental and functional state of the individual responding T cell.

Finally, we examined the relationship between anti-CD4 mAb-induced effects on CD3 and CD4 surface expression. Cells that had most reduced their surface CD4 expression in response to anti-CD4 cross-linking displayed the greatest increase in surface CD3 expression, whereas cells that had least reduced their surface CD4 expression displayed no change in surface CD3 expression (Table 1, experiment 2). Thus, diminished CD4 expression and enhanced CD3/TCR expression seem to reflect two related responses of double positive thymocytes to antibody-induced CD4 cross-linking.

Because the natural ligand of CD4 is probably intrathymic Ia (refs 8-13), physiological CD4/Ia interactions might be expected to increase TCR expression on double positive cells in a manner analogous to CD4/anti-CD4 interactions. In fact, a few double positive cells in a normal thymus do express high levels of the TCR^{14,15}, possibly because these cells express

Fig. 3 Effects of short-term anti-CD4 mAb treatment on CD4⁺8⁺ and CD4⁺8⁻ thymocytes. Thymocytes from mice injected for two days with anti-CD4 mAb or saline were stained and analysed by two-colour flow cytometry. Two-parameter data files were software-gated to generate single-colour staining histograms of CD4 (a and c) and CD3 (b and d) on CD8⁺ (predominately double positive) thymocytes (a and b) and on CD8⁻ (CD4⁺8⁻ single positive and CD4⁻8⁻ double negative) thymocytes (c and d). Each single-colour histogram shows the indicated stain for the gated thymus cell population from control saline-injected mice (solid line, designated 'C'), and the indicated stain for the gated thymus cell population from experimental anti-CD4 mAb-injected mice (solid line, designated 'I'). The relative staining intensities of cells from anti-CD4 mAb-injected mice compared to cells from control mice are as follows: a, 31%; b, 315%; c, 8%; d, 69%. Two-colour contour representations are shown for CD8 versus CD4 staining (e and f) and CD8 versus CD3 staining (g and h), and indicate by the horizontal line the division used for software-gating to select CD8⁺ cells and CD8⁻ cells, which was based on negative control antibody staining.

Methods. C3H pregnant female mice were purchased from the Frederick Cancer Research Facility, Frederick, Maryland. Starting on day 8 after birth, neonates were injected intraperitoneally for two days with 200 µg purified GK1.5 anti-CD4 mAb, or with saline. On day 10, thymocytes were prepared and stained for CD4, CD8 and CD3, and then analysed by two-colour flow cytometry. CD4 staining was with GK1.5 culture supernatant followed by MAR18.5-FITC, CD8 staining was with 53-6.72-Biotin plus Texas Red streptavidin, and CD3 staining was with 145-2C11-FITC. Negative control stains were Leu4-FITC and Leu1-Biotin. Fluorescence data were collected using logarithmic amplification on 5 × 10⁴ viable cells, as determined by forward light scatter and propidium iodide exclusion. Identification of CD8⁺ and CD8⁻ thymocytes was based on Texas Red (CD8) staining, and software-gating was then used to generate the FITC staining histograms (CD4 and CD3) of those two populations. For quantitative comparisons, median channel values of fluorescence intensity of the relevant cell populations were converted to linear units using empirically derived standard calibration curves.

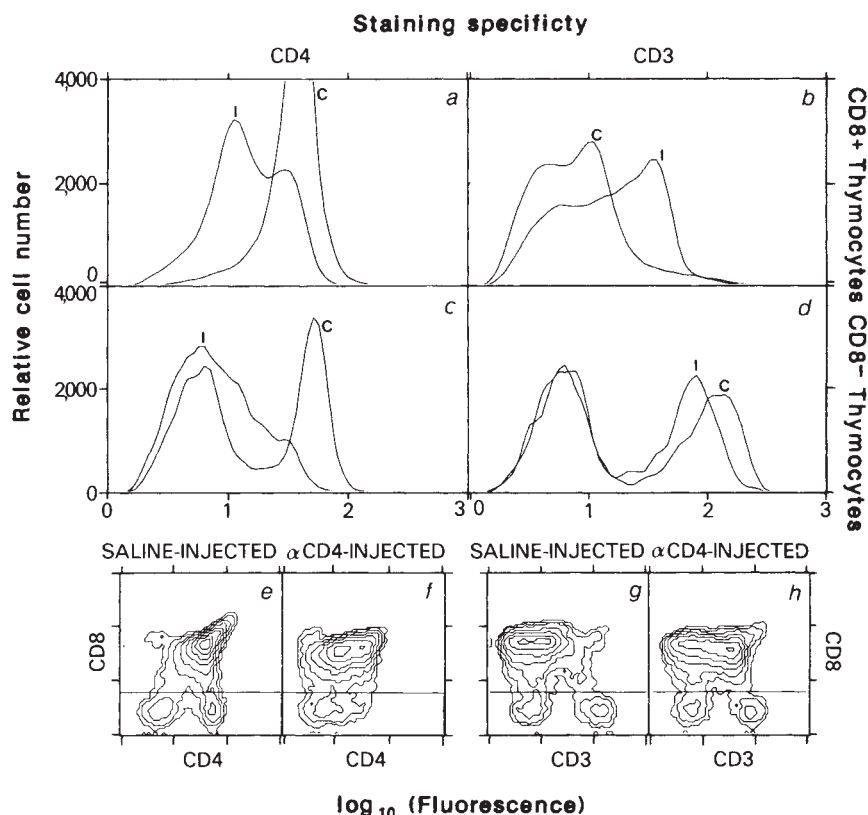


Table 1 Effect of anti-CD4 mAb on surface expression of CD3 and CD4

Experiment	Group	In vivo treatment mAb (duration)	Thymocyte population	% Control expression	
				CD3	CD4
1	1	Saline control	CD8 ⁺	100	100
	2	Anti-CD4 (1 day)	CD8 ⁺	206	45
	3	Anti-CD4 (2 days)	CD8 ⁺	318	34
	4	Anti-CD4 (3 days)	CD8 ⁺	316	29
2	5	Saline control	Total	100	100
	6	Anti-CD4 (13 days)	Total	233	21
	6	Anti-CD4 (13 days)	CD4-dull	313	15
	6	Anti-CD4 (13 days)	CD4-bright	100	61

In experiment 1, C3H mice were injected with 100 µg GK1.5 anti-CD4 mAb intraperitoneally for one day (day 9 after birth), two days (days 8 and 9 after birth) or three days (days 7, 8 and 9 after birth). On day 10, thymocytes from the control and three experimental groups were stained and analysed by two-colour flow cytometry for CD8 versus CD3 staining and for CD8 versus CD4 staining. In experiment 2, C3H mice were injected with 100 µg GK1.5 intraperitoneally per day on the first seven days after birth, and 200 µg GK1.5 intraperitoneally per day on days 8-13 after birth. On day 14, thymocytes from the control and experimental groups were stained and analysed by two-colour flow cytometry for CD3 versus CD4 staining. The indicated thymocyte populations were identified and characterized by software gating of flow cytometry data. In experiment 1, CD8⁺ cells were identified as those cells with CD8 staining intensities above the negative control mAb. It should be noted that in such 10-day-old thymuses >98% of CD8⁺ cells are CD4⁺8⁺ cells and <2% are CD4⁻8⁺ cells. In experiment 2, CD4-dull cells from anti-CD4 mAb-injected mice were identified as those cells with CD4 staining intensities below 90% of the control thymocyte population; this represented 74% of the cells from the mAb-injected population. Conversely, CD4-bright cells were identified as those cells from the anti-CD4 mAb-injected thymocyte population with CD4 staining above that cut-off; this represented the remaining 26% of the cells from the mAb-injected population. 'Total' indicates that no sub-populations were defined, and that all viable cells were used in the data analysis. The relative staining intensities were calculated as described in the Fig. 1 legend, and then compared with the saline-injected control population, defined as 100% for each experiment.

Ia-specific antigen receptors able to promote productive CD4/Ia interactions. Thus, CD4/Ia interactions, in addition to contributing to the overall avidity of T-cell interactions with Ia, may also promote T-cell differentiation by increasing TCR expression on developing double positive thymocytes. From this perspective, the inductive effect on double positive thymocytes of anti-CD4, but not anti-Ia, mAbs can be explained if the anti-CD4 mAb partly mimics normal CD4/Ia interactions whereas anti-Ia mAb blocks normal CD4/Ia interactions¹⁶⁻¹⁸.

We are grateful to Dr Randy Hardy, Fox Chase Cancer Centre, Philadelphia, and Ms Julie Titus, NIH, for conjugated antibodies; and to Drs William Biddison, Richard Klausner, Jean Maguire, Elizabeth Shores and Dinah Singer for critically reading this manuscript.

Received 19 August; accepted 28 September 1988.

1. Fowlkes, B. J., Edison, E., Mathieson, B. J. & Chused, T. M. *J. exp. Med.* **162**, 802-888 (1985).
2. Reinherz, E. L. & Schlossman, S. F. *Cell* **19**, 821-827 (1980).
3. Shortman, K. & Jackson, H. *Cell. Immun.* **12**, 230-246 (1974).
4. McPhee, D., Pye, J. & Shortman, K. *Thymus* **1**, 151-162 (1979).
5. Havran, W. L. *et al. Nature* **330**, 170-173 (1987).
6. Finkel, T. H., McDuffie, M., Kappler, J. W., Marrack, P. & Cambier, J. C. *Nature* **330**, 179-181 (1987).
7. McDuffie, M., Born, W., Marrack, P. & Kappler, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8728-8732 (1986).
8. Biddison, W. E., Rao, P. A., Talle, M. A., Goldstein, G. & Shaw, S. *J. exp. Med.* **156**, 1065-1076 (1982).
9. Swain, S. *Immunol. Rev.* **74**, 129-142 (1983).
10. MacDonald, H. R., Glasebrook, A. L., Bron, C., Kelso, A. & Cerottini, J.-C. *Immunol. Rev.* **68**, 89-115 (1982).
11. Greenstein, J. L., Kappler, J., Marrack, P. & Burakoff, S. J. *J. exp. Med.* **159**, 1213-1224 (1984).
12. Golding, H. *et al. Nature* **317**, 425-428 (1985).
13. Doyle, C. & Strominger, J. L. *Nature* **330**, 256-259 (1987).
14. Bluestone, J. A., Pardoll, D., Sharrow, S. O. & Fowlkes, B. J. *Nature* **326**, 82-84 (1987).
15. Crispe, I. N., Shimonkevitz, R. P., Husmann, L. A., Kimura, J. & Allison, J. P. *J. Immun.* **139**, 3585-3589 (1987).
16. Kruisbeek, A. M., Fultz, M. J., Sharrow, S. O., Singer, A. & Mond, J. J. *J. exp. Med.* **157**, 1932-1946 (1982).
17. Kruisbeek, A. M. *et al. J. exp. Med.* **161**, 1029-1047 (1985).
18. Mizuochi, T., Tentori, L., Sharrow, S. O., Kruisbeek, A. M. & Singer, A. *J. exp. Med.* **168**, 437-442 (1988).
19. Dyalinas, D. P. *et al. J. Immun.* **131**, 2445-2451 (1983).
20. Leo, O., Foo, M., Sachs, D. H., Samelson, L. E. & Bluestone, J. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1374-1378 (1987).
21. Staerz, U. D., Rammensee, H.-G., Benedetto, J. D. & Bevan, M. J. *J. Immun.* **134**, 3994-4000 (1985).
22. Lanier, L. L., Gutman, G. A., Lewis, D. E., Griswold, S. T. & Warner, N. L. *Hybridoma* **1**, 125-131 (1982).
23. Sarmiento, M., Glasebrook, A. L. & Fitch, F. W. *J. Immun.* **125**, 2665-2672 (1980).
24. Janeway, C. A. Jr *et al. J. Immun.* **132**, 662-667 (1984).
25. Pierres, A. *et al. J. Immun.* **132**, 2275-2285 (1984).

Cell-cell adhesion mediated by CD8 and MHC class I molecules

Anne M. Norment*, Russell D. Salter†, Peter Parham†, Victor H. Engelhardt‡ & Dan R. Littman*§

* Department of Microbiology and Immunology, and
§ Howard Hughes Medical Institute, University of California at
San Francisco, California 94143, USA

† Department of Cell Biology, Stanford University, Stanford,
California 94305, USA

‡ Department of Microbiology, University of Virginia, Charlottesville,
Virginia 22908, USA

CD4 and CD8 are cell-surface glycoproteins expressed on mutually exclusive subsets of peripheral T cells. T cells that express CD4 have T-cell antigen receptors that are specific for antigens presented by major histocompatibility complex class II molecules, whereas T cells that express CD8 have receptors specific for antigens presented by MHC class I molecules (reviewed in ref. 1). Based on this correlation and on the observation that anti-CD4 and anti-CD8 antibodies inhibit T-cell function, it has been suggested that CD4 and CD8 increase the avidity of T cells for their targets by binding to MHC class II or MHC class I molecules respectively^{2,3}. Also, CD4 and CD8 may become physically associated with the T-cell antigen receptor, forming a higher-affinity complex for antigen and MHC molecules⁴⁻⁸, and could be involved in signal transduction^{5,6,9,10}. Cell-cell adhesion dependent CD4 and MHC II molecules has recently been demonstrated¹¹. To determine whether CD8 can interact with MHC class I molecules in the absence of the T-cell antigen receptor, we have developed a cell-cell binding assay that measures adhesion of human B-cell lines expressing MHC class I molecules to transfected cells expressing high levels of human CD8. In this system, CD8 and class I molecules mediate cell-cell adhesion, showing that CD8 directly binds to MHC class I molecules.

We anticipated that the binding between CD8 and MHC class I molecules might be a low-affinity interaction, similar to that of other cellular adhesion receptors^{12,13}. Because detection of such interactions usually requires multivalent binding, stable cell lines with a high degree of CD8 expression were developed (Fig. 1). The relative levels of CD8 expression by the four CHO lines are shown in Fig. 1a.

Surface MHC class I expression of several B lymphoblastoid lines used in the cell-cell binding assay was determined using the monoclonal antibody (mAb) W6/32 (Fig. 1b) which is specific for monomorphic determinants of class I molecules¹⁴. The B cells included Raji, RM3 and CIR-A2.1, all of which express high levels of MHC class I (10^6 - 10^7 molecules per cell, data not shown), and Daudi and CIR, which express significantly reduced levels of surface class I molecules. Daudi cells are deficient in class I because of an absence of β_2 -microglobulin expression. CIR cells¹⁵ lack the *HLA-A* and *HLA-B* genes, but nonetheless retain significant binding activity for mAb W6/32, probably because of its reactivity with the *HLA-C* locus product or with the recently described non-classical class I molecules¹⁶. CIR cells did not react with the HLA-A2, HLA-A28-specific mAb CR11-351 (ref. 17) (Fig. 1b). CIR-A2.1, a cell line obtained by transfecting an HLA-A2.1 genomic clone¹⁸ into CIR, expressed levels of class I molecules comparable with Raji.

A cell-cell binding assay was developed for the study of the interaction between CD8 and MHC class I molecules in the absence of the T-cell antigen receptor (TCR). As shown in Fig. 2, MHC class I⁺ B cells (Raji, RM3, CIR-A2.1) specifically bound to CD8⁺ CHO.3 and CHO.4 cells at levels 15-30-fold greater than to CD8⁻ CHO.1 cells. As many as 90% of the class I⁺ B cells remained tightly bound to CD8⁺ CHO cells and were not removed from the plates with further washing. The extent of cell-cell binding correlated with the relative level of CD8 expression on the surface of the CHO cells (Fig. 2). MHC class I⁺ B cells failed to bind to the CD8⁻ line CHO.1 and to CHO.2, which expressed levels of CD8 at least five-fold higher than those on peripheral T cells. Binding was detected only when a threshold level of CD8 expression was achieved. Thus CHO.3 which expressed ~8-fold more surface CD8 than CHO.2 bound class I⁺ B cells, and CHO.4, which expressed double the level of CD8 compared with CHO.3, bound even more effectively. Cell-cell binding also correlated with the relative number of class I molecules on the B cells. RM3 (ref. 19) expressed twice as many class I molecules as Raji cells (Fig. 1b), and bound the most effectively. CIR did not specifically bind to CD8⁺ CHO cells. It is possible that the W6/32-reactive molecules expressed by CIR (HLA-C or non-classical class I molecules) may not bind to CD8, or that their expression level may be insufficient to mediate cell-cell adhesion.

The CD8-MHC class I mediated cell-cell binding was specifically and completely inhibited by anti-MHC I and anti-

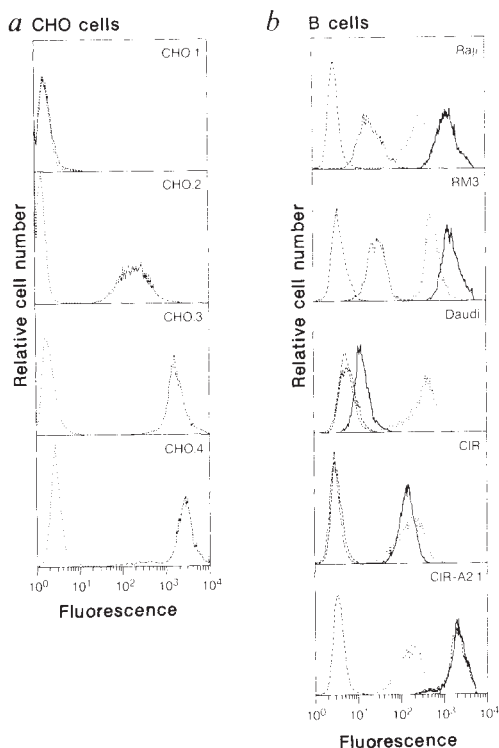
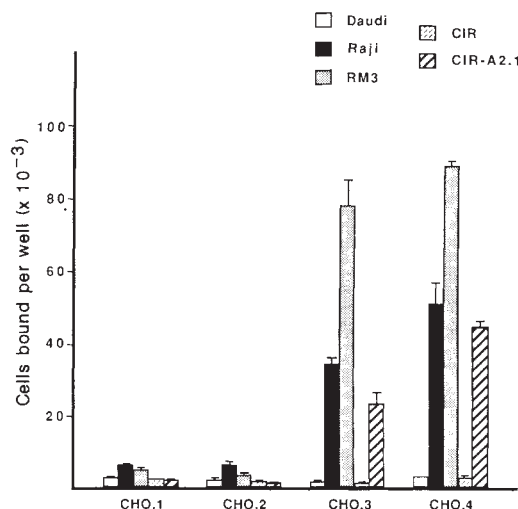


Fig. 1 Cytofluorometric analysis of CHO- and B-cell lines. Cells were incubated with saturating levels of the anti-human transferrin receptor mAb L01.1 (dotted line), the anti-CD8 α mAb OKT8 (dotted-dashed line), W6/32 (ref. 14) (solid line), or the HLA-A2, HLA-A28-specific mAb Cr11-351 (ref. 17) (dashed line). Staining used a second step fluorescein-conjugated goat anti-mouse IgG antibody (Becton Dickinson). Fluorescence was measured in arbitrary units using a FACScan cell analyser (Becton Dickinson). **a**, CHO-cell lines. CHO.1 (CD8 $^{-}$) and CHO.2 (CD8 $^{+}$) were derived by selection in media lacking hypoxanthine. CHO.3 and CHO.4 were obtained by methotrexate amplification of CHO.2. **b**, B-cell lines. Raji is a MHC class I $^{+}$ Burkitt lymphoma³¹. RM3 (ref. 19) (from M. Peterlin) is a class II $^{-}$ derivative of Raji. Daudi is Burkitt lymphoma deficient in surface MHC I as a result of a defect in β_2 -microglobulin³². CIR (Hmy 2 CIR) is an HLA-A, B-negative derivative of Licrlon Hmy 2 (ref. 15). CIR-A2.1 (ref. 18) was obtained by transfection of CIR with an HLA-A2.1 genomic clone under the control of its endogenous promoter.

Methods. The CHO-DHFR $^{-}$ cell line DXB11 (cell culture facility, UCSF) was co-transfected by calcium phosphate precipitation³³ with the CD8 α F1.1 complementary DNA (ref. 24) inserted into the vector pSV7d, a derivative of pHS210 (ref. 34) (from P. Luciw), and the plasmid pMG1P (ref. 35) bearing the hamster DHFR minigene (from D. L. Chasin). Primary selection was in Ham's F12 lacking hypoxanthine as described³⁶. Colonies were screened for the expression of CD8 by rosetting²⁴, then expanded and rescreened by immunofluorescence analysis. A negative (CHO.1) and a positive (CHO.2) clone were chosen. CHO.3 was obtained by selection of CHO.2 in 0.03 μ M methotrexate (Sigma). CHO.4 is a pool of colonies obtained by selection of CHO.2 in 0.06 μ M methotrexate.

Fig. 2 Binding of B cells to CHO cells expressing different levels of CD8. B-cell lines are indicated by the key at the top of the figure: MHC I $^{+}$ lines (Raji, RM3, CIR-A2.1), MHC I $^{-}$ lines (Daudi, CIR). CHO lines are indicated across the base of the histogram: CHO.1 is CD8 $^{-}$; CHO.2, CHO.3 and CHO.4 express increasing levels of CD8. The number of cells bound per well was determined. Error bars represent standard deviations of triplicate samples.

Methods. Cell-cell binding was assayed in flat-bottom 96-well microtitre dishes (Corning) in phosphate buffered saline (PBS) with 10% fetal bovine serum. B cells were radiolabelled for 2 h with [35 S]cysteine (12.5 μ Ci per 10^6 cells, >600 Ci mmol $^{-1}$, Amersham). 10^5 B cells were added to wells, each of which contained an adherent confluent monolayer of about 10^5 CHO cells plated two days earlier. The B cells were centrifuged (160g) for 5 min on to the CHO cells, and the plates incubated for 1 h at 37 $^{\circ}$ C on an orbital shaker. Wells were washed gently 10 times with 10% fetal bovine serum/PBS; the cells were solubilized in 1% Nonidet P-40 (Sigma), 400 mM NaCl in PBS, and the amount of radioactivity of the lysate was determined by scintillation counting. The number of cells bound per well was determined on the basis of the known specific activity of the B cells.



CD8 mAbs (Fig. 3). W6/32 blocked binding of Raji and CIR-A2.1 to the CD8 $^{+}$ CHO.4 cells (Fig. 3a). The HLA-A2, HLA-A28-specific mAb CR11-351 did not block binding of Raji cells, which have only slight reactivity with the antibody and express HLA-A1 and A3 (A. Calman, personal communication), but completely blocked binding of CIR cells transfected with HLA-A2.1. The anti-human transferrin receptor mAb L01.1 was reactive with the B-cell lines, but did not block specific binding. Ca206 (ref. 20), a mAb reactive with monomorphic determinants of human HLA-D molecules, also did not block binding (data not shown). Specific cell-cell binding was also completely inhibited by anti-CD8 mAbs that recognize distinct epitopes of human CD8 (refs 21, 22) (Fig. 3b).

These experiments demonstrate that the human CD8 glyco-

protein directly binds to MHC class I molecules, supporting the idea that CD8 is an adhesion receptor that increases the avidity of T cells for their targets. Although we have detected binding in the absence of the TCR, this binding was of low avidity; over-expression of CD8 was required to mediate stable adhesion between transfected cells and B-cell lymphoblastoid lines. We cannot exclude the possibility that other molecules on either cell surface contribute to the specific cell-cell adhesion in our assay system. It has been shown, however, that beads coated with lipid bilayers bearing purified class I molecules can effectively stimulate CD8 $^{+}$ T cells in the absence of any other molecules²³.

The binding experiments we describe were performed with only the CD8 α chain, which forms homodimers and homomul-

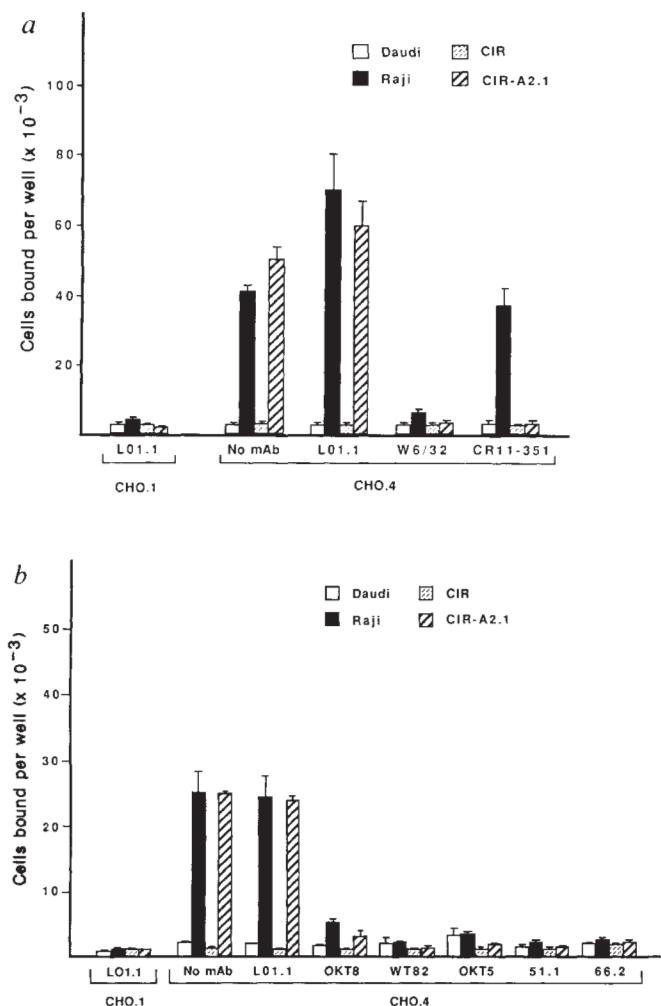


Fig. 3 Inhibition of CD8-MHC I binding by anti-MHC I and anti-CD8 mAbs. B-cell lines are as indicated in Fig. 2. Binding to CHO.1(CD8⁻) or CHO.4(CD8⁺) was determined. *a*, Inhibition by anti-MHC I mAb. B cells were preincubated with saturating levels (determined by fluorescence-activated cell sorting) of the following mAbs: L01.1 (anti-human transferrin receptor), W6/32 (anti-MHC I monomorphic) or CR11-351 (anti-HLA-A2, HLA-A28-specific). Control cells not incubated with a mAb are also indicated (no mAb). *b*, Inhibition by anti-CD8 mAb. CHO cells were preincubated with saturating levels of L01.1 or one of several anti-CD8 mAbs (OKT8, WT82, OKT5, 51.1, 66.2). Control cells (no mAb) are indicated. B cells were pre-incubated with saturating levels of normal mouse serum (Xymed) to prevent binding of anti-CD8 mAb by Fc receptors.

Methods. B cells or CHO cells were incubated with the indicated mAb for 30 min at 4 °C in 0.01% sodium azide, 5% fetal calf serum in PBS, and then washed with 10% fetal bovine serum in PBS. For B cells, incubation used 10 µg antibody per 10⁶ cells in 0.1 ml. CHO cells were pre-incubated with 2.5 µg antibody per well in 0.1 ml. Three anti-CD8 mAbs (WT82, 51.1, 66.2) completely blocked binding; two mAbs (OKT8 and OKT5) inhibited 90% of specific binding. The cell-cell binding assay was performed as described in the legend to Fig. 2, except that 0.01% sodium azide was added. Antibodies included in these studies were affinity-purified by protein A Sepharose chromatography, except for OKT5, which was in the form of ascitic fluid and used at a dilution of 1:25. The source of each antibody is as follows: L01.1 (Becton Dickinson); WT82 (Dr W. Tax, University of Nijmegen); OKT5 (Dr E. Reinherz, Dana-Farber Cancer Institute); 51.1 (ref. 21) and 66.2 (ref. 21) (Dr P. Martin, F. Hutchinson Cancer Center).

timers in transfected cells²⁴. Experiments with T-cell hybridomas indicate that a CD8-dependent response to antigen can be reconstituted with the TCR and CD8 α -chain alone²⁵. Rodent CD8, however, is expressed primarily as a heterodimer composed of the Lyt2 (α) and Lyt3 (β) chains²⁶, and we have recently shown that a CD8 β (Lyt3) protein is expressed on human CD8⁺ T cells²⁷. Whether CD8 β has a role in class I adhesion and/or signal transduction is not yet known.

Binding has now been observed between transfected cells over expressing CD4 or CD8 and lymphocytes expressing MHC class II (ref. 11) or class I molecules. As these interactions are of low avidity, it is unlikely that CD4 or CD8 alone can initiate cell-cell adhesion, assuming no regulated increase in their binding affinities. It is more likely that CD4 and CD8 stabilize cell-cell binding, probably in cooperation with other T-cell adhesion receptors and with the TCR. CD4 and CD8 could also facilitate TCR-mediated signal transduction by binding to a limited number of MHC molecules after the T cell is already bound to its target. This is supported by the demonstration of synergistic activation of T cells by anti-CD4 or anti-CD8 mAbs and anti-CD3 mAbs^{5,6,9,10}. Whether CD4 and CD8 serve to cluster MHC molecules in the vicinity of TCRs and/or effect active signalling remains to be resolved.

It has been proposed that CD8, like the TCR, binds to the α -helices formed by the α_1 and α_2 -domains of class I molecules²⁸⁻³⁰. Binding of CD8 to this site would preclude simultaneous interaction of the TCR and CD8 with a single class I molecule. The CD8-class I cell-cell binding assay can be used to map amino-acid residues of class I involved in binding to CD8 and to determine whether CD8 and the TCR bind to the same or different regions of the class I molecule.

We thank L. Chasin for the DHFR vector, A. Calman and M. Peterlin for cell lines, W. Tax, P. Martin and E. Reinherz for mAbs, and A. DeFranco and A. Weiss for reading the manuscript. This research was supported by a grant from the NIH (D.R.L.), a Searle Scholar award (D.R.L.), and a Medical Scientist Training Program Fellowship (A.M.N.).

Received 15 July; accepted 22 September 1988.

- Littman, D. R. *Rev. Immun.* **5**, 561-584 (1987).
- Swain, S. L. *Immun. Rev.* **74**, 129-142 (1983).
- Marrack, P. *et al. J. exp. med.* **158**, 1077-1091 (1983).
- Saizawa, K., Rojo, J. & Janeway Jr, C. A. *Nature* **328**, 260-263 (1987).
- Takada, S. & Engleman, E. G. *J. Immun.* **139**, 3231-3235 (1987).
- Rivas, A., Takada, S., Koide, J., Sonderstrup-McDevitt, G. & Engleman, E. J. *Immun.* **140**, 2912-2918 (1988).
- Weyand, C. M., Goronzy, J. & Fathman, G. J. *Immun.* **138**, 1351-1354 (1987).
- Anderson, P., Blue, M. L. & Schlossman, S. F. *J. Immun.* **140**, 1732-1737 (1988).
- Emmrich, F., Kanz, L. & Eichmann, K. *Eur. J. Immun.* **17**, 529-534 (1987).
- Ledbetter, J. A., June, C. H., Grosmaire, L. S. & Rabinovitch, P. S. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1384-1388 (1987).
- Doyle, C. & Strominger, J. L. *Nature* **330**, 256-259 (1987).
- Pierschbacher, M. D. & Ruoslahti, E. *Nature* **309**, 30-33 (1984).
- Selvaraj, P. *et al. Nature* **326**, 400-403 (1987).
- Barnstable, C. J. *et al. Cell* **14**, 9-20 (1978).
- Storkus, W. J., Howell, D. N., Salter, R. D., Dawson, J. R. & Cresswell, P. J. *Immun.* **138**, 1657-1659 (1987).
- Geraghty, D. E., Koller, B. H. & Orr, H. T. *Proc. natn. Acad. Sci. U.S.A.* **84**, 9145-9149 (1987).
- Russo, C., Ng, A. K., Pellegrino, M. A. & Soldano, F. *Immunogenetics* **18**, 23-35 (1983).
- Hogan, K. T. *et al. J. Immun.* **141**, 2519-2535 (1988).
- Calman, A. F. & Peterlin, B. M. *J. Immun.* **139**, 2489-2495 (1987).
- Charron, D. J. & McDevitt, H. O. *J. exp. Med.* **152**, 188-366 (1980).
- Martin, P. J., Ledbetter, J. A., Clark, E. A., Beatty, P. G. & Hansen, J. A. *J. Immun.* **132**, 759-765 (1984).
- Biddison, W. E., Rao, P. E., Talle, M. A., Boselli, C. M. & Goldstein, G. *Human Immun.* **9**, 117-130 (1984).
- Goldstein, S. A. N. & Mescher, M. F. *J. Immun.* **140**, 3707-3711 (1988).
- Littman, D. R., Thomas, Y., Maddon, P. J., Chess, L. & Axel, R. *Cell* **40**, 237-246 (1985).
- Gabert, J. *et al. Cell* **50**, 545-554 (1987).
- Ledbetter, J. A. & Seaman, W. E. *Immun. Rev.* **68**, 197-218 (1982).
- Norment, A. M. & Littman, D. R. *EMBO J.* (in the press).
- Bjorkman, P. J. *et al. Nature* **329**, 506-512 (1987).
- Bjorkman, P. J. *et al. Nature* **329**, 512-518 (1987).
- Davis, M. M. & Bjorkman, P. J. *Nature* **334**, 395-402 (1988).
- Pulvertaft, R. J. V. *Lancet* **i**, 238-240 (1964).
- De Prével, C. & Mach, B. *Immunogenetics* **17**, 133-140 (1983).
- Graham, F. L. & Van der Eb, A. J. *Virology* **52**, 456-467 (1973).
- Stuvé, L. L. *et al. J. Virol.* **61**, 326-335 (1987).
- Mitchell, P. J. *et al. Molec. cell. Biol.* **6**, 425-440 (1986).
- Nunberg, J. H., Kaufman, R. J., Schimke, R. T., Urlaub, G. & Chasin, L. A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5553-5556 (1978).

A non-cell-autonomous mutation regulating juvenility in maize

Scott Poethig

Plant Science Institute, Department of Biology,
University of Pennsylvania, Philadelphia,
Pennsylvania 19104-6018, USA

The transition from a juvenile to an adult phase of shoot growth in plants involves coordinated changes in a large number of morphological and physiological traits, including the ability of the plant to undergo sexual reproduction. Although the environmental, nutritional and chemical factors that influence this process (phase change) have been intensively studied for many years, its basic mechanism is still unknown^{1,2}. An analysis of the genetic regulation of phase change has been facilitated by the identification of several mutations of maize (*Teopod1*, *Tp2*, *Tp3* and *Corngrass*) that prolong the expression of a juvenile developmental programme^{3,4}. All of these mutations have a striking atavistic phenotype, the most dramatic feature of which is the transformation of reproductive structures into leaves or leaf-like structures. Here I report that X-ray-induced sectors expressing the wild-type allele of *Tp1* in otherwise mutant plants possess a mutant rather than a wild-type phenotype. This non-cell-autonomous behaviour is highly unusual and suggests that *Tp1* controls the production or distribution of a diffusible factor(s) that regulates juvenile development.

Tp1 is a semi-dominant, gain-of-function mutation with a highly pleiotropic phenotype characterized by the following features: an increase in the number of vegetative segments and axillary branches; a decrease in the size of leaves and internodes; an increase in the number of leaves with epidermal wax and internodes with prop roots; a reduction in the size and degree of branching of the ear and tassel; and the transformation of reproductive organs into leaves or leaf-like structures⁴ (Fig. 1). *Tp1* plants containing sectors of wild-type tissue were created by irradiating seeds heterozygous for *Tp1* (*o5 Tp1/o5-1241 Tp1*⁺). Losses of *Tp1* were marked by the expression of *o5-1241*, a chlorophyll mutation located approximately 10 map units proximal to *Tp1*. Because terminal deletions or somatic recombination events that uncover *o5-1241* necessarily uncover loci distal to this gene, cells expressing *o5-1241* were likely to be hemizygous or homozygous for the wild-type allele of *Tp1*. Most X-ray-induced aberrations in maize seem to be terminal deletions or large interstitial deficiencies⁵; consequently, a majority of the sectors observed were probably hemizygous for this wild-type allele. These sectors were expected to be fully viable because plants hemizygous for the entire region of the chromosome distal to *o5* can be produced using B-A translocations and are morphologically normal, although somewhat smaller than normal⁶. Seeds were X-irradiated (1 krad, 250 kV, 15 mA, 2 mm Al) 24 hours after imbibition, grown to maturity, and then screened for sectors expressing the yellow-to-white phenotype of *o5-1241*. The genotype of the epidermis overlying a sector was determined by examining guard cells with fluorescence microscopy; guard cells possessing chlorophyll produce red fluorescence when excited with blue light, whereas albino guard cells only exhibit the pale green fluorescence of the cell wall^{7,8}. I concentrated on the reproductive phenotype of *Tp1*, in particular the presence of leaves or leaf-like structures in the ear or tassel, because this is the most distinctive feature of this mutation.

Twenty-six tassel sectors, 14 ear sectors and 15 tiller sectors were produced in a total of 3,500 plants. Thirty-two of these sectors originated in the L1 (epidermal) layer of the meristem and 23 in the L2 (sub-epidermal) layer. Most of the structures within a sector contained both *Tp1* (green) and normal (white) tissues, as would be expected because lateral structures arise

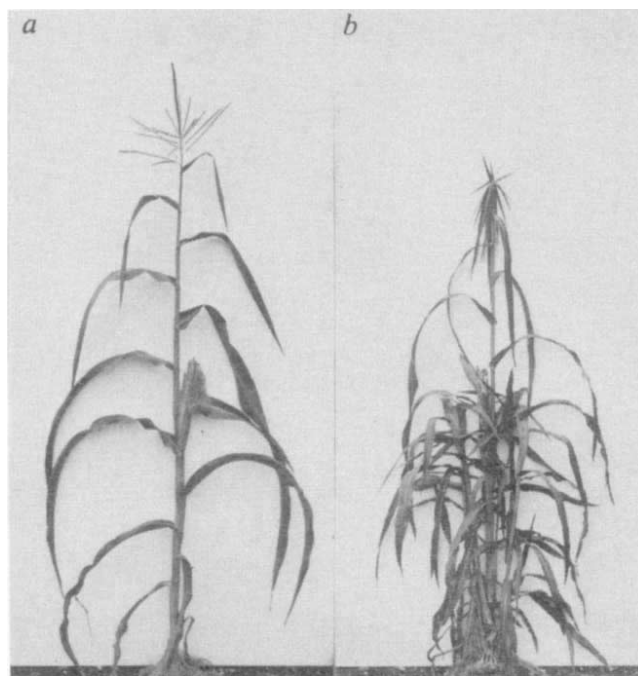


Fig. 1 Wild-type (a) and *Tp1/Tp1*⁺ (b) maize plants.

from several layers of the shoot apical meristem⁹⁻¹¹. In several cases, however, I observed genetically homogeneous structures in regions of a sector in which *o5-1241* cells in the L1 had displaced underlying L2 cells (Fig. 2b). Apart from three sectors confined to the tassel, all of the sectors present in the ear and the tassel were also present in adjacent vegetative segments of the shoot. Tassel sectors encompassed an average of 5.1 vegetative nodes, whereas tiller sectors always extended over the entire length of the tiller. Sectors occupied 1/20 to 1/10 of the circumference of the tassel, but occupied a much larger fraction—often the entire circumference—of an ear or a tiller. This difference reflects the fact that ear and tiller buds are derived from a fraction of cells in the circumference of the meristem, whereas the tassel samples its entire circumference^{9,11}.

With one exception, all of the sectors had a mutant phenotype indistinguishable from that of adjacent *Tp1/Tp1*⁺ tissue, even when the structures within the sector consisted entirely, or almost entirely, of wild-type tissue (Fig. 2). The exception was a tassel sector that produced a completely albino lateral branch with normal spikelets. But, although this case is particularly striking, it is unclear whether to attribute the normal phenotype of this lateral branch to its genotype or to the poor expressivity of *Tp1*, because other non-sectored parts of the tassel had a normal phenotype.

One possible explanation for the lack of concordance between the genotype and phenotype of these sectors is that *Tp1* acts extremely early in shoot development and is not required after this stage. Thus, loss of *Tp1* at a seedling stage of development is irrelevant to the fate of meristem cells because that fate has already been fixed. An alternative possibility is that *Tp1* controls the production or distribution of a diffusible factor that regulates the fate of both mutant and normal cells. In this case, the continued expression of a mutant phenotype after the loss of *Tp1* would be due to the diffusion of such a factor from surrounding mutant tissue into wild-type cells within the sector.

Although our results do not allow us to distinguish unequivocally between these possibilities, we prefer the latter explanation because there is no evidence that the fate of shoot meristem cells is determined early in shoot development. In particular, the fact that almost all the sectors observed in this study encompassed both vegetative and reproductive structures demonstrates

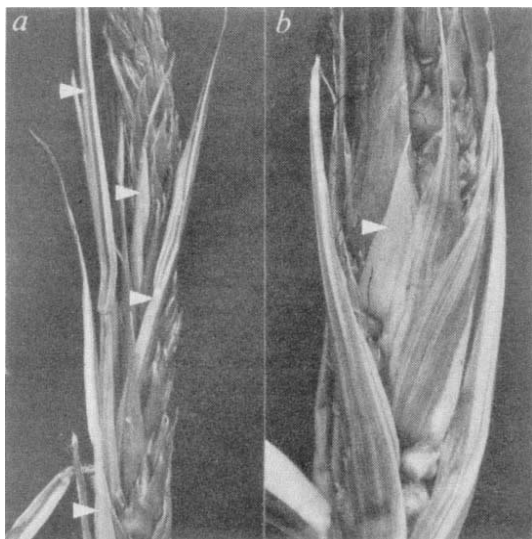


Fig. 2 The morphology of X-ray-induced *o5-1241 Tp1*⁺ (yellow green) sectors in plants of the genotype *o5 Tp1/o5-1241 Tp1*⁺. Note that structures within these sectors express a mutant phenotype indistinguishable from that of structures outside the sectors. *a*, A tassel sector located in the L2 lineage of the shoot meristem. *o5-1241 Tp1*⁺ cells are present in most of the internal tissue of the leaves within the sector (arrows), but are not present in the epidermis and in green portions of the leaf, which are derived from the (green) epidermis. *b*, An L1 sector in an ear-like portion of a tassel. *o5-1241 Tp1*⁺ cells are present in the epidermis and in the marginal mesophyll of the leaves within the sector. One leaf within the sector is entirely *o5-1241 Tp1*⁺ (arrow).

that the determination of reproductive lineages occurred after the time of irradiation; similar results have been obtained with wild-type maize irradiated at the time of germination¹¹. Thus the expression of a *Tp1* phenotype by wild-type sectors cannot represent the maintenance of an abnormal floral programme established early in shoot development. It therefore seems that wild-type sectors express a mutant phenotype because of the activity of the surrounding mutant cells.

Non-cell-autonomous genes are rare in maize. All of the 40 or so genes whose expression has been examined in genetic mosaics are cell-autonomous or almost completely so; in the few exceptional cases, non-autonomy is only observed over a distance of one or a few cells^{8,9,12,13}. Of the morphological mutations that have been examined to date, only *Knotted* (*Kn*) is expressed non-cell-autonomously⁸. In this case, mutant mesophyll tissue is capable of inducing aberrant cell divisions in overlying wild-type epidermal cells, but *Kn* epidermal cells have no effect on the behaviour of the mesophyll. The high degree of non-autonomy observed in the case of *Tp1* is therefore fairly unusual, and may reflect a unique regulatory function.

The possibility that *Tp1* controls the production or distribution of a diffusible factor is consistent with reports that diffusible factor(s) are important in the maintenance of the juvenile state in woody species¹⁵⁻¹⁹. Although a number of studies have implicated the hormone gibberellic acid in this process, the evidence for this is still inconclusive¹. Chemical characterization of the product of *Tp1* should therefore add significantly to our understanding of the regulation of juvenile development in plants.

I wish to thank J. Nadeau, K. Poethig and A. Bassiri for technical assistance, and A. Cashmore, F. Daldal and W. Telfer for critical comments on the manuscript. This work was supported by the NSF.

Received 1 July; accepted 29 September 1988.

- Zimmerman, R. H., Hackett, W. P. & Pharis, R. P. in *Encyclopedia of Plant Physiology N.S.* Vol. 11 (eds Pharis, R. P. & Reid, D. M.) 79-115 (Springer, Berlin, 1985).
- Allsopp, A. in *Encyclopedia of Plant Physiology* Vol. 15 (ed. Ruhland, W.) 1172-1221 (Springer, Berlin, 1965).

- Galinat, W. *Maize Genet. Coop. News Lett.* **40**, 102-103 (1966).
- Poethig, R. S. *Genetics* **119**, 959-973 (1988).
- Fabergé, A. C. Z. *indukt. Abstram.-W. Vererblehre* **87**, 392-420 (1956).
- Chang, M. T., Coe, E. H. & Beckett, J. B. *Maize Genet. Coop. News Lett.* **61**, 91-92 (1987).
- Stewart, R. N. & Dermen, H. *Am. J. Bot.* **66**, 47-58 (1979).
- Hake, S. & Freeling, M. *Nature* **320**, 621-632 (1986).
- Johri, M. M. & Coe, E. H. *Dev. Biol.* **97**, 154-172 (1983).
- Poethig, R. S. in *Contemporary Problems in Plant Anatomy* (eds White, R. A. & Dickison, W. C.) 235-259 (Academic, New York, 1984).
- McDaniel, C. N. & Poethig, R. S. *Planta* **175**, 13-22 (1988).
- Bennetzen, J. L., Blevins, W. E. & Ellingboe, A. H. *Science* **241**, 208-210 (1988).
- Coe, E. H. & M. G. Neuffer in *Corn and Corn Improvement* (ed. Sprague, G. F.) 111-223 (Am. Soc. Agron., Madison, 1976).
- Freeling, M. & Hake, S. *Genetics* **111**, 617-634 (1985).
- Doorenbos, J. *Proc. K. ned. Akad. Wet. Series C* **57**, 99-102 (1954).
- Stoutemyer, V. T. & Britt, O. K. *Nature* **189**, 854-855 (1961).
- Frank, H. & Renner, O. *Planta* **47**, 105-114 (1956).
- Schwabe, W. W. & Al-Doori, A. H. *J. exp. Biol.* **24**, 969-981 (1973).
- McDaniel, C. N. *Planta* **148**, 462-467 (1980).

The catalytic subunit of cAMP-dependent protein kinase induces expression of genes containing cAMP-responsive enhancer elements

Karl T. Riabowol*, J. Stephen Fink†||, Michael Z. Gilman*, Donal A. Walsh‡, Richard H. Goodman† & James R. Feramisco§

* Cold Spring Harbor Laboratory, Box 100, Cold Spring Harbor, New York 11724, USA

† Molecular Medicine, Tufts-New England Medical Center, Boston, Massachusetts 02111, USA

‡ Department of Biological Chemistry, University of California at Davis, Davis, California 95616, USA

§ Cancer Center, University of California at San Diego, 225 Dickinson St, San Diego, California 92103, USA

Transcriptional regulation of eukaryotic genes by cyclic AMP requires a cAMP-dependent protein kinase (A kinase)¹. Two hypotheses have been proposed to explain how the holoenzyme of the A kinase induces transcription. The regulatory subunits of the A kinase, which bind cAMP² and DNA³, and have amino-acid homology with the *Escherichia coli* catabolite activator protein⁴ could directly stimulate gene expression^{5,6}. Alternatively, phosphorylation by the catalytic subunits could induce transcription by activating proteins involved in gene transcription⁷⁻⁹. To distinguish between these models, we microinjected purified preparations of the catalytic and regulatory subunits of A kinase into tissue culture cells and monitored expression of a stably integrated fusion gene containing a cAMP-responsive human promoter¹⁰ fused to a bacterial reporter gene, or of the endogenous *c-fos* gene. The catalytic subunit stimulated expression of these genes, whereas the regulatory subunit did not. These results indicate that the catalytic subunit of A kinase is sufficient to induce expression of two cAMP-responsive genes, without increasing levels of cAMP.

In *E. coli*, cAMP regulates the expression of genes involved in sugar catabolism by binding to catabolite activator protein. Subsequent binding of the cAMP-CAP complex to specific DNA sequences activates transcription of cAMP-regulated genes¹¹, in contrast to the eukaryotic mechanism involving A kinase¹². The holoenzyme of A kinase exists as an enzymatically inactive tetramer, composed of two catalytic (C) and two regulatory (R) subunits¹³. When levels of cAMP are elevated in eukaryotic cells, cAMP binds to the R subunits and the holoenzyme dissociates, yielding an R-subunit dimer and two active C subunits. It

|| Present address: Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

Fig. 1 Sequences necessary for transcriptional induction by cAMP. *a*, The VIP- β -gal fusion gene which was co-transfected into C6 glioma cells with pRSV-neo^r (a plasmid containing a neomycin resistance gene expressed from the Rous sarcoma virus promoter/enhancer) contains 2,000 base pairs (bp) of 5' untranslated sequences of the human VIP gene and 146 bp of the VIP untranslated exon 1, fused to the bacterial β -galactosidase reporter gene. Inverted repeats within the VIP CRE are shown by wavy arrows. The CRE in the VIP gene is located between 70 and 86 bp upstream from the transcriptional initiation site¹⁰. This 17-bp sequence is necessary for transcriptional regulation of the VIP gene by cAMP, as mutations within the -C-G-T-C-A- motifs of this region abolish cAMP-induced expression of VIP-CAT fusion genes²⁵. The level of induction of the intact gene (a fusion gene containing 2,000 bp of the VIP 5' flanking region linked to CAT) and a fusion gene containing the VIP CRE fused to a viral promoter are equivalent, indicating that the CRE is entirely responsible for cAMP-induction of VIP gene transcription²⁵. *b*, The 5' region of the mouse *c-fos* gene contains a region of dyad symmetry called the serum response element (SRE). Two regions showing identity with a consensus CRE are indicated by wavy arrows. The sequence located between -71 and -59 has been shown to contain a cAMP-responsive element (L. A. Berkowitz and M.Z.G., manuscript in preparation).

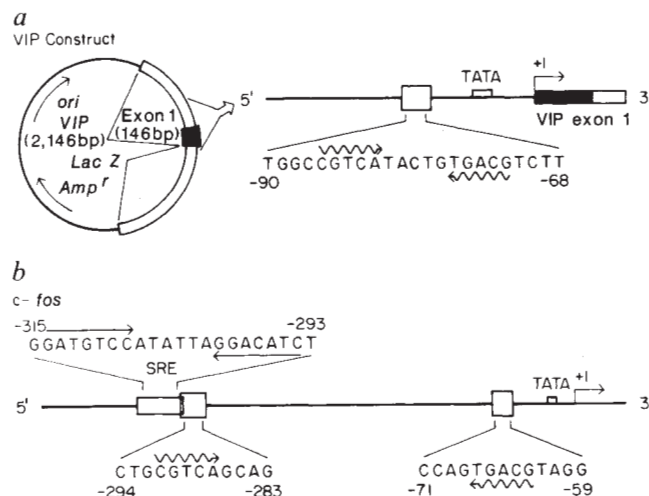
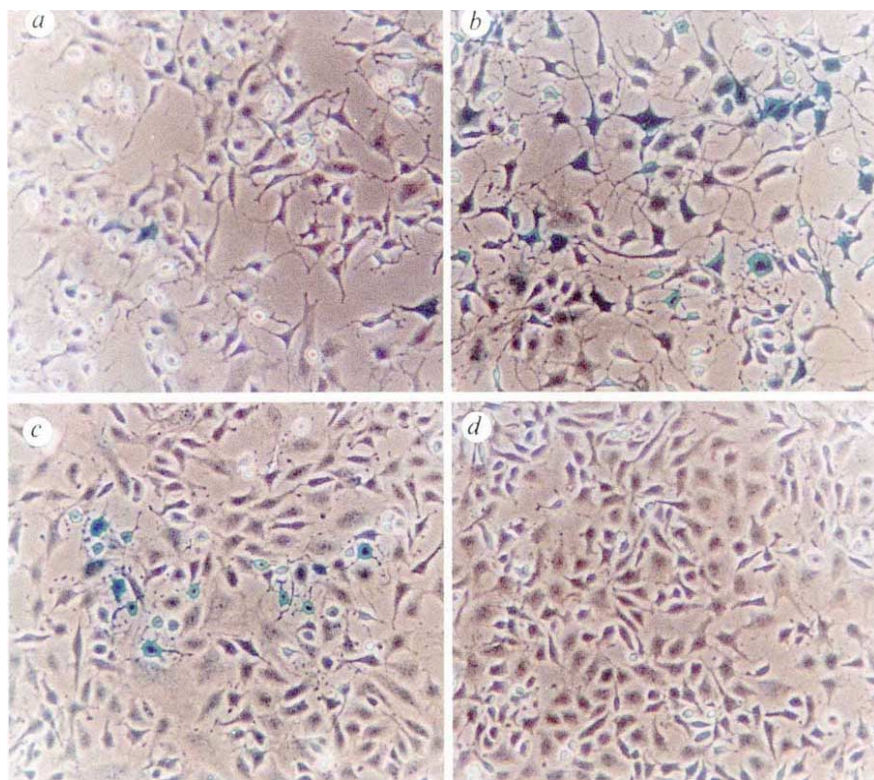


Fig. 2 Microinjection of C subunit of cAMP-dependent protein kinase induces expression of a VIP- β -galactosidase fusion gene. *a*, *b*, Effect on β -galactosidase expression of incubating cells with IBMX and forskolin for 3 and 6 h, respectively. The same proportion of cells stained for the presence of β -galactosidase if treated for 6, 12 or 24 h. *c*, Cells microinjected with C subunit of A kinase, fixed and stained to detect β -galactosidase activity 3 h after injection. No cells injected with equal amounts of C and R₁₁ subunits expressed β -galactosidase (*d*).

Methods. Rat C6 glioma cells were grown as fibroblasts¹⁹. Cells were microinjected using glass capillary needles and, unless otherwise indicated, R₁, R₁₁ and C subunits of A kinase were injected at 0.5 mg ml⁻¹. This concentration of A kinase subunit elevates endogenous levels of this protein 2–4-fold²². Intracellular levels of cAMP were elevated in C6 glioma cells by treating cells with 5 μ M forskolin (an activator of adenyl cyclase) in ethanol plus 0.5 mM IBMX in dimethylformamide. Dimethylformamide was added in controls. Cells treated with drugs or microinjected with purified proteins were incubated at 37 °C for the indicated times and fixed for 10 min at room temperature in phosphate-buffered saline (PBS) containing 2% formaldehyde and 0.2% glutaraldehyde. Following extensive rinsing with PBS, expression of β -galactosidase was assayed by incubation of cells in PBS containing 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide, 2 mM magnesium chloride and 0.2 mg ml⁻¹ of X-gal for 24 h at 37 °C.



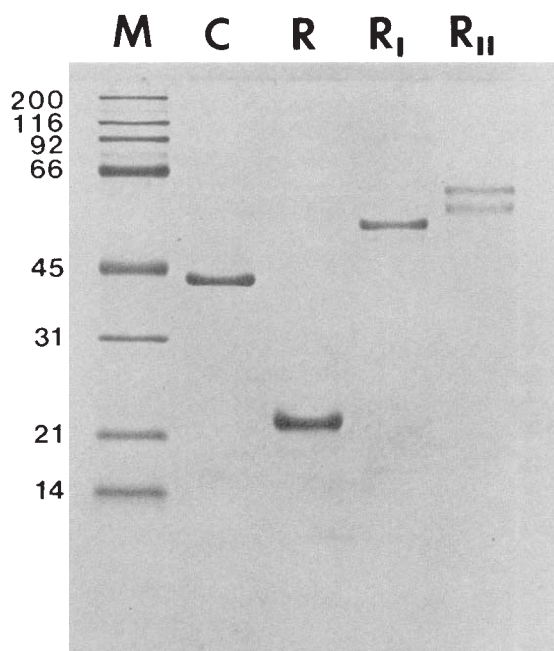
has been suggested that both the C and R subunits of A kinase mediate the transcriptional regulation of cAMP-responsive genes. The R subunits could function in a manner similar to bacterial CAP, and upon binding cAMP, act directly upon target DNA sites in cAMP-responsive genes⁶. In support of this hypothesis, cAMP-responsive elements (CREs) in some eukaryotic genes are homologous to the bacterial CAP-binding sequences. In particular, the -C-G-T-C-A- motif, essential for activity of CREs, is also present within the CAP-binding site consensus sequence¹⁰.

Alternatively, the C subunit of A kinase could alter gene expression by phosphorylation of chromosomal or other nuclear

proteins^{7–9}. This hypothesis is supported by several observations. First, the C subunit migrates from the cytoplasm into the nucleus after treatment of cells with agents that increase intracellular levels of cAMP¹⁴. Second, co-transfection of tissue culture cells with a construct encoding an active protein-kinase inhibitor peptide¹⁵, decreased the cAMP-induced expression of a human proenkephalin-bacterial chloramphenicol acetyl transferase (CAT) fusion gene¹⁶. Finally, the cAMP-regulated expression of endogenous hepatic genes was blocked by treating cells with three different inhibitors of A kinase¹⁷. Although these studies indicate that the active C subunit of A kinase is required for expression of cAMP-responsive genes, it is possible that a

Fig. 3 Proteins microinjected. Aliquots of ras (R), A kinase catalytic subunit (C), R_I and R_{II} were boiled in Laemmli sample buffer and electrophoresed through a 10% polyacrylamide gel containing sodium dodecyl sulphate. After staining of the gel with Coomassie blue dye, two bands were observed for the R_{II} subunit, one of which represents a phosphorylated form of the protein. Lane M shows protein standards (BioRad) of the indicated relative molecular mass.

Methods. The catalytic and R_{II} subunits of A kinase were purified from bovine heart²³. The R_I subunit and ras (T-24; cloned from a bladder carcinoma containing a glycine to valine substitution at residue 12) proteins were over-expressed and purified from bacteria. Purified proteins were stored on ice as stock solutions containing 10 mM Tris (pH 7.0), 100 mM NaCl, 0.1 mM EDTA and 1 mM β -mercaptoethanol. Immediately before use they were diluted to the desired concentration in the same buffer but without EDTA and β -mercaptoethanol.



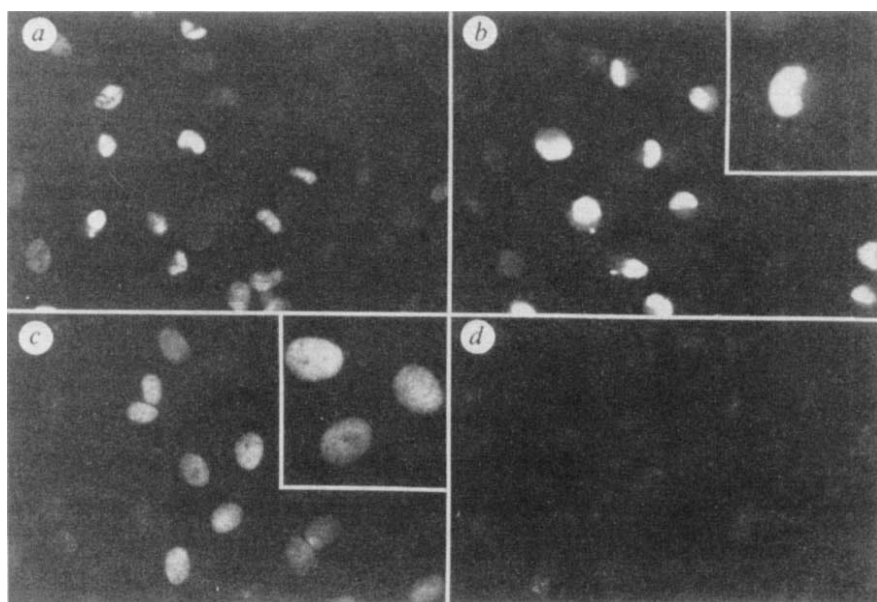
cAMP-binding protein is also required for response, or that phosphorylated R subunits are involved in transcriptional activation by cAMP.

To distinguish between these two models, we microinjected purified C and R subunits, separately and together, and determined their effects on the expression of cAMP-regulated genes. Two different assays were used to measure cAMP-responsive gene expression. First, we established a C6 glioma cell line containing a stably integrated fusion gene consisting of the human vasoactive intestinal peptide (VIP) gene promoter/enhancer region joined to the *E. coli lac Z* gene (Fig. 1a). Expression of the fusion gene in individual cells was detected by staining using the chromogenic β -galactosidase substrate X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactoside). Treatment with 5 μ M forskolin and 0.5 mM isobutyl methylxanthine (IBMX, a phosphodiesterase inhibitor) for 3 h (Fig. 2a), or 6 h (Fig. 2b), induced β -galactosidase expression as detected by the appearance of cells that stain blue with X-gal. Identical results were obtained when cells were treated with 0.5 mM 8-bromo cAMP. At the cell density used in these experiments, approximately 50% of the cells expressed β -galactosidase after treatment with forskolin and IBMX or with 8-bromo cAMP.

To determine the role of the C and R subunits of A kinase in expression of the fusion gene, purified subunit proteins were microinjected into the glioma cells. Aliquots of each of the protein preparations used are shown in Fig. 3. Injection of the purified C subunit resulted in the expression of β -galactosidase in the injected cells (Fig. 2c). The level of expression was much greater than that obtained when cells were treated with IBMX and forskolin for the same length of time (compare Fig. 2a and c). Regulation of the VIP-*lac Z* fusion gene was also examined in cells injected with type I or type II regulatory subunits (R_I or R_{II}), or with the C subunit plus varying amounts of the R_{II} subunit. No cells were stained when equal amounts of C and R_{II} subunits were injected (Fig. 2d). In all experiments in which buffer R_I or R_{II} alone, or a greater amount of R_{II} subunit than

Fig. 4 Microinjection of the catalytic subunit of cAMP-dependent protein kinase induces expression of *c-fos*. Logarithmically growing rat 208F fibroblasts²⁴ were made quiescent by incubation in medium containing 0.1% fetal calf serum for 24 h before injection. The purified C subunit of A kinase was injected and the cells were incubated for 90 min (a) or 3 h (b) before fixation and staining for the fos protein. Ras protein (c), and a mixture of equal amounts of C and R_{II} subunits (d) were also injected, and the cells were fixed and processed for immunocytochemistry with fos antibody following 3 h (ras) or 6 h (C + R_{II}) of incubation at 37 °C. No nuclear fluorescence was observed in cells injected with buffer, or with the mixture of C and R_{II} when cells were fixed and processed for immunofluorescence 1, 2 or 4 h after injection (data not shown). Insets in b and c show higher magnification micrographs of the nuclei of C subunit and ras injected cells, respectively.

Methods. 208F fibroblasts were grown using the culture conditions described in Fig. 2. Unless otherwise indicated, ras protein (T-24) was injected at a concentration of 1 mg ml⁻¹. For the determination of fos expression, cells were microinjected, incubated for various lengths of time, and fixed by sequential immersion for 10 min in PBS containing 3.7% formaldehyde, 0.5% Triton X-100 and 0.05% Tween 20. Immunofluorescent detection of fos protein was done by incubating fixed cells for 1 h at 37 °C, in PBS containing 1% bovine serum albumin and 50 μ g ml⁻¹ of affinity-purified rabbit anti-fos antibody, followed by incubation with fluorescein-conjugated goat anti-rabbit immunoglobulin G as described¹⁹.



of the C subunit were injected, no expression of β -galactosidase was observed.

As a second assay, we wished to examine the regulation of an endogenous gene within the context of its native chromatin configuration. Thus, we measured the appearance of fos protein in the nuclei of cells microinjected with the different subunits of A kinase. Although the *c-fos* DNA sequences responsible for regulation by cAMP have not been fully defined, several -C-G-T-C-A- motifs are found in sequences flanking the gene (Fig. 1b). At least one of these sequences, which binds a protein factor and contributes to basal promoter activity¹⁸, is c-AMP responsive (L. A. Berkowitz and M.Z.G., manuscript in preparation). Quiescent rat fibroblasts were microinjected with the C subunit, ras protein, or mixtures of C and R_{II} subunits. Expression of the fos protein was monitored by indirect immunofluorescence using affinity-purified antibodies raised against the v-fos protein¹⁹. By 90 min, expression of the fos protein was markedly induced in cells injected with the C subunit (Fig. 4a), and by 3 h very high levels of nuclear fos staining were observed (Fig. 4b). A mixture of equal amounts of C and R_{II} subunits failed to induce fos expression in these cells (Fig. 4d). Injection of ras protein, which is also known to induce fos expression in fibroblasts²⁰, was included as a positive control (Fig. 4c). Injection of C subunit induced expression of the fos protein even though little if any induction was seen when cells of the same line were treated with 8-bromo cAMP or with forskolin and IBMX. Similarly, injection of the C subunit, but not treatment with 8-bromo cAMP or forskolin and IBMX²¹ induced c-fos expression in NIH 3T3 cells (data not shown).

These experiments show that the C subunit of A kinase is sufficient to induce the expression of at least two genes that are regulated by cAMP, and indicate that protein phosphorylation is involved in this process. The experiments in which a mixture of C and R subunits were injected indicate that R_{II} can inactivate the C subunit, probably by forming a catalytically inactive tetramer. Furthermore, the fact that R_I or R_{II} were unable to induce transcription by themselves or when coinjected with cAMP (not shown), implies that these regulatory subunits are not crucial in the expression of these genes. Because these microinjection experiments were performed in the absence of agents that increase intracellular levels of cAMP, they also suggest that activation of cellular proteins by the binding of cAMP is not required for expression of the genes tested.

In addition to demonstrating that active catalytic subunit of cAMP-dependent protein kinase is sufficient to induce the expression of genes known to be transcriptionally activated by elevated levels of cAMP, these studies describe a method for visualizing the activation of a eukaryotic enhancer at the single-cell level. Injection of purified proteins or antibodies into appropriate indicator cells should provide a powerful method for evaluating the role of putative transcription factors in the regulation of genes through different signal transduction pathways.

We thank Philip Renna for photographic services, Christine Arrigo for technical assistance and Dr Susan Taylor for kindly supplying the R_I subunit of A kinase. This work was supported by USA NIH-PHS grants (to M.Z.G., J.R.F. and R.H.G.). K.T.R. was supported by an American Heart Association (Nassau County Branch) fellowship and an NIH training grant.

Received 23 August; accepted 22 September 1988.

- Montminy, M. R., Sevarino, K. A., Wagner, J. A., Mandel, G. & Goodman, R. H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6682-6686 (1986).
- Robinson-Steiner, A. M. & Corbin, J. D. *J. biol. Chem.* **258**, 1032-1039 (1983).
- Jungmann, R. A., Constantinou, A. I., Squinto, S. P., Kwant-Welfeld, J. & Schweppe, J. S. *Ann. N.Y. Acad. Sci.* **478**, 147-161 (1986).
- Weber, I. T., Takio, K., Titani, K. & Steitz, J. A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7679-7683 (1982).
- Constantinou, A. I., Squinto, S. P. & Jungmann, R. A. *Cell* **42**, 429-437 (1985).
- Nagamine, Y. & Reich, E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4606-4610 (1985).
- Waterman, M., Murdoch, G. H., Evans, R. M. & Rosenfeld, M. G. *Science* **229**, 267-269 (1985).
- Montminy, M. R. & Bilezikian, L. M. *Nature* **328**, 175-178 (1987).

- Yamamoto, K. K., Gonzales, G. A., Biggs III, W. H. & Montminy, M. R. *Nature* **335**, 494-498 (1988).
- Tsukada, T., Fink, J. S., Mandel, G. & Goodman, R. H. *J. biol. Chem.* **262**, 8743-8747 (1987).
- DeCrombrugge, B., Busby, S. & Buc, H. *Science* **224**, 831-838 (1984).
- Beavo, J. A., Bechtel, P. J. & Krebs, E. G. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3580-3583 (1974).
- Edelman, A. M., Blumenthal, D. K. & Krebs, E. G. *Rev. Biochem.* **56**, 567-613 (1987).
- Nigg, E. A., Hiltz, H., Eppenberger, H. M. & Dutly, F. *EMBO J.* **4**, 2801-2806 (1985).
- Cheng, H. C., Van Patten, S. M., Smith, A. J. & Walsh, D. A. *Biochem. J.* **231**, 655-661 (1985).
- Grove, J. R., Price, D. J., Goodman, H. M. & Avruch, J. *Science* **238**, 530-533 (1987).
- Buchler, W., Walter, U., Jastorff, B. & Lohmann, S. M. *FEBS Lett.* **228**, 27-32 (1988).
- Gilman, M. Z., Wilson, R. N. & Weinberg, R. A. *Molec. cell Biol.* **6**, 4305-4316 (1986).
- Riabowol, K. T., Vosatka, R. J., Ziff, E. B., Lamb, N. J. & Feramisco, J. R. *Molec. cell Biol.* **8**, 1670-1676 (1988).
- Stacey, D. W., Watson, T., Kung, H. F. & Curran, T. *Molec. cell Biol.* **7**, 523-527 (1987).
- Bravo, R. *et al. Cell* **48**, 251-260 (1987).
- Lamb, N. J. C. *et al. J. Cell Biol.* **106**, 1955-1971 (1988).
- Fletcher, W. H., Van Patten, S. M., Chen, H. C. & Walsh, D. A. *J. biol. Chem.* **261**, 5504-5513 (1986).
- Quade, K. *Virology* **98**, 461-466 (1979).
- Fink, J. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 6662-6666 (1988).

Squid major lens polypeptides are homologous to glutathione S-transferases subunits

Stanislav I. Tomarev & Rina D. Zinovieva

N. K. Koltzov Institute of Developmental Biology, USSR Academy of Sciences, 26 Vavilov Street, 117808 Moscow, USSR

The eye lenses of cephalopods and vertebrates evolved relatively recently and by independent routes¹. They provide a good experimental model for the study of convergent evolution at the protein level. One proposal is that pre-existing proteins were recruited as structural eye lens proteins during evolution². This has been confirmed for the vertebrate eye lens structural proteins, or crystallins³⁻⁹, which have been intensively studied^{10,11}. Despite the limited information about cephalopod eye lenses¹²⁻¹⁴, it has been suggested that glutathione S-transferases (GSTs) are a possible evolutionary ancestor of the squid major lens proteins. Recently, the N-terminal sequence of the squid major lens protein was shown³ to be 55% homologous with that of the Y_a subunit of the rat GST. Here, we demonstrate that the squid major lens polypeptides are encoded by a gene family of at least three members. We characterize two cDNAs corresponding to these genes and show they probably either are GST subunits themselves, or share an evolutionary ancestor with them.

To investigate the squid major lens polypeptides (SMLPs) and their evolutionary relationship with GSTs we have obtained a complementary DNA library from the poly(A)⁺ RNA of the lens of the squid *Ommastrephes sloanei pacificus*. We isolated the recombinant clones which gave the strongest hybridization

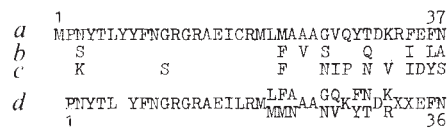


Fig. 1 Comparison of the N-terminal region of the SMLPs encoded by pSI10, pSI17, pSI20 and pSI21 cDNAs (a); pSI11 cDNA (b); pSI12 and pSI18 cDNAs (c) with the N-terminal region of the squid *N. gouldi* major protein¹² (d). There are nine polymorphisms in the *N. gouldi* protein and both residues are shown at these positions.

Methods. cDNA was synthesized²³ using poly(A)⁺ RNA from the squid lens as a template, size-fractionated by agarose electrophoresis and fractions longer than 600 base pairs were cloned into an *Sma* site of pUC19 plasmid. Sequencing was performed by combination of the Maxam-Gilbert technique²⁴ and a dideoxy method using covalently-closed plasmid DNA as template²⁵.

Fig. 2 Nucleotide sequences of cDNAs corresponding to clones pSI20 (I), pSI11 (II) and their deduced amino-acid sequences (III, IV). Sequences corresponding to pSI20 clone are written in full. In the sequences corresponding to the pSI11 clone, only differing positions are shown. An asterisk marks the start of the pSI11 sequence. Dashes indicate gaps which were introduced to maximize homology.

-58	TTCAATCAGTACTCTCCAGATAACCTCTGTGTTCACACTTCTCTGAACTCAGT		
		*AGTAAAT CT CTT CA AC	
1	ATGCCTAACTACACTTTGCTACTACTTCAATGGCCCGTGGACGTGCGGAGATCTGCGGTATGTGTGATGGCCGCGTGTCCAGTATAC	I	
	TC CC G G T C C T C T T TTC T A CAG	II	
1	MetProAsnTyrThrLeuTyrTyrPheAsnGlyArgGlyArgAlaGluIleCysArgMetLeuMetAlaAlaGlyValGlnThr	III	
	Ser Phe Val Ser Gln	IV	
91	GATAAGAGCGTTCGAATTCACGAATGGGACAAATACAGAACGATATGCTAGCATGTGCTGCTCCCTGTTTGGACATGTATGGCCGAGAAC	I	
	C A T AGC AC C G T A CTA A AT CATATGC A C G AC G C	II	
31	AspLysArgPheGluPheAsnGluTrpAspLysTyrArgAsnAspMetProSerMetCysValProValLeuAspIleAspGlyGlnAsn	III	
	Ile LeuAla ThrGlnPheLysThrLys CysHisMetLeu Ile Glu ThrGluThr	IV	
181	AAAAATGCCCGAAACAATGGCCATTTGCCAGTACTTGGCCAGAGAAAACGGTTACTACGGCAAGAACACATGGACATGTTCAGAAATCGAT	I	
	C GG T C G GT T A T G TT T T T A T T AGG T	II	
61	LysMetProGluThrMetAlaIleAlaArgTyrLeuAlaArgGluAsnGlyTyrTyrGlyLysAsnAsnMetMetPheArgIleAsp	III	
	GlnVal GlnSer Ser Phe Phe	IV	
271	TACATCTGTGATTGCTTCTATGAATCCTCCAGATTACATAGGTTACTTCCATACTAAGAACGGTTCATGCGAGGGAAGCGGTACT	I	
	G C C C A C TC T C T A	II	
91	TyrIleCysAspCysPheTyrGluIleLeuHisAspTyrMetArgTyrPheHisThrLysAsnGlyArgPheMetGlnGlySerGlyThr	III	
	CysLeu SerLeuPhe LeuPheAsn	IV	
361	GATATGAGCCCTGACATGGACCCCAACAGATGACATCTTACATTCAGAAATAGATATTTGGATACCTGTGCTAGGATCCTTCTCTCTTG	I	
	CCG TATAAC A A G TG A A TGAG TG--- A G TCCAAA C T AG AC T ATA	II	
121	AspMetSerProAspMetAspProThrGlnMetThrSerTyrIleGlnAsnArgTyrLeuAspThrCysLeuArgIleLeuSerPheLeu	III	
	AlaValTyrAsnGluLys AlaAlaLysLys GluLeu - Lys PheGlnAsn Val ProTyrMet	IV	
451	GAAAGAACCCCTTGAATGGTAAATGGTGGAAAGAGTTCTTCATGGGTGACAGATGATGTTGTGTGATATGATGCTGCTATTTGCTGCTTG	I	
	AG G GCTAAC G A TGCC GC GG C T TC TC C AC C GC GC C	II	
151	GluArgThrLeuGluMetArgAsnGlyLysGluPheMetGlyAspGlnMetMetCysTyrCysCysLeu	III	
	Lys AlaAsnLys AlaGlyTrp Ile IleLeu ThrHisAlaAla	IV	
541	GAGAACCCCTTGTGGAGACACAGACCAATTCACAACTTCCCAAACTGTATGTCATGTGGTAACAGCTGTGCTTCCCACTCCATAGATC	I	
	CCAA AA GCC ATCTC G GG A AT G AGCCG TC A CC C AG T C A T	II	
181	GluAsnProMetLeuGluAspGlnThrThrPheAsnAsnPheProLysLeuMetSerLeuTrpLysArgValAlaSerHisProLysIle	III	
	IleGln AsnAlaAsnLeuLeuLysGluTyr AlaAla ArgThr Ala	IV	
631	ACCCCTTATCTGAAAAAGCGTAACAAACCACTGGTATGATACCCAACTATATGCCCAAGGTGAATACAATGGTCCCTATGAAAG	I	
	G TG G CA C G A TGCT TC AGC GTAG CTTCCT GGAATACT TAACGA GAGGCCAGTAGA TTA	II	
211	ThrProTyrLeuLysLysArgAsnAsnThrAsnTrpTer	III	
	AlaAla Ile AlaPhe	IV	
721	AATCTGATTGTGATATCGGCCCTTTTATCAAGGATGACTGGATGATTTCTGGAAGTGAAGAAATTTTATTACTTTGACTCATTTATTT	I	
	TA ATT ATTG GACA TAAGGCGGTT TGTC A A CAT C AAA CCT GTCTCTACGC CC GGGAGG CG TGC T A	II	
811	TGTATCAAACTGATGATCTTAAATAATATCTCTTTTAAAT(A) ₁₈	I	
	CTACCATCTTTAT AATATA AC CGA G AACGTTGAATGATTTTATATTTGTAATGGACCAACGCAGCTTTTCTTTTCAAAA	II	
901	CAAAATGCGATTGTTTGATATATCAATAAAATTTTATGTAAGCCCTT(A) ₁₅	II	

signals to a [³²P]cDNA probe synthesized from squid lens poly(A)⁺ RNA.

At least seven of the 18 clones selected code for polypeptides whose N-terminal sequences are 80–94% homologous with that of the major lens protein of the squid *Nototodarous gouldi*¹² (Fig. 1). The complete nucleotide and deduced amino-acid sequences of two of these clones (pSI20 and pSI11) are shown in Fig. 2. The calculated relative molecular masses of the polypeptides they encode are 26,200 and 23,700 respectively. The apparent relative molecular masses of the SMLPs are between 27,000 (27K) and 30,000 (30K) (ref. 3; R.D.Z., K. S. Aleinikova, A. T. Mikhailov and S.I.T., manuscript in preparation). On the basis of these results we concluded that cDNAs of clones pSI20, pSI11 and pSI12 are representatives of a gene family, of at least three members, coding for SMLPs.

There is 60% homology between cDNAs of clones pSI20 and pSI11 in the coding region. The residues coding for the N-terminal regions of SMLPs are more highly conserved than those coding for the C-terminal regions; moreover, the regions coding for the first 20 amino acid residues are the most highly conserved (88% homology). There is no homology in the 5' and 3' non-translated regions. At the amino acid level, there is 50% homology between the 23,700 and 26,200 polypeptides.

The 26,200 polypeptide has an extremely high methionine content (about 10%) as do the major lens protein of *N. gouldi*¹² and *Loligo pealii*¹³ (9.6–12.8%) and also the fish γ -crystallins (13–14%)¹⁵. The methionine content of the 23,700 polypeptide is lower (about 5%), indicating that in the squid major lens proteins, which are mainly dimers of subunits¹², the 26,200 polypeptide is more abundant than the 23,700 molecule. Of the 18 clones selected, four contained cDNA inserts for the pSI20 clone, whereas there is only one copy of the cDNA insert of the pSI11 clone and two of the pSI12 clone.

The squid major lens protein may be related to the rat GST Y_a subunit³. GSTs are a multifunctional family of dimeric proteins present in most cells. They are detoxification enzymes important for protection against peroxidative damage^{16–18}. GST

subunits probably have a common evolutionary origin despite differences in amino-acid sequences^{19,20}. There are 12 invariant residues and conservative changes in 24 positions²¹. A comparison of the amino-acid sequences of the SMLPs with sequences of rat GST Y_a and Y_b subunits using the DIAGON program²² (window = 15, score = 170) showed that the regions of homology were residues 1–32, 65–81, 159–174 and 194–214 (Fig. 3). No similarity with vertebrate crystallins was identified by these criteria. Only 20–25% of the amino-acid residues are identical in SMLPs and GST subunits, but the 12 invariant residues of the GST subunits are conserved in SMLPs except for a threonine residue at position 58 in the 23,700 polypeptide. In addition, 20 residues have conservative changes only, out of the 24 in SMLPs. At the nucleotide level there is 36–38% homology in the coding regions of mRNAs for SMLPs and rat Y_a and Y_b subunits.

On the basis of our results and those of Wistow and Piatigorsky³, we propose that SMLPs share an evolutionary ancestor with GST subunits or are GST subunits themselves. Low levels of GST activity in lens extracts³ may be due to modification of SMLPs with age or to the presence of inhibitory factors³. Investigation of authentic GST or other squid tissues and the expression of SMLP cDNAs in appropriate tissues may reveal whether SMLPs are really GST subunits.

Although the region in GSTs between amino acids 50 and 85 is thought to be important for glutathione binding or recognition of hydrophobic substrates²¹, the most highly conserved region within the SMLPs and compared with GST subunits is between amino acids 1 and 27. This N-terminal region is probably also important in GST molecules, for example in the formation of dimers by subunits.

It seems that the formation of the squid lens occurred mainly by recruitment of proteins of one class (GST subunits), but other types of proteins may have been recruited in other cephalopod lenses. We have shown that tubulins are among the major polypeptides of octopus lens (R.D.Z., K. S. Aleinikova, A. T. Mikhailov and S.I.T., manuscript in preparation). Pre-

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

Imaging plate illuminates many fields

Y. Amemiya and J. Miyahara

The erasable phosphor imaging plate developed for medical radiography has found new uses in the laboratory. X-ray diffraction, protein crystallography and autoradiography have all benefited from this technology transfer from the clinic.

SEVERAL years ago, researchers at Fuji Film developed an erasable X-ray imaging plate for diagnostic radiography, based on the X-ray excitation of a phosphor layer¹. Recently, the imaging plate has been used to obtain quantitative intensity measurements in X-ray diffraction experiments^{2,3} and in autoradiography^{4,5}. The results show that the imaging plate is much more useful than conventional X-ray film, establishing it as a rival to other X-ray image sensors such as X-ray television cameras, solid-state image devices and gas-type area detectors.

The imaging plate is approximately 0.5 mm in thickness, and is composed of a flexible plastic plate coated with fine photostimulable phosphor crystals (BaFBr:Eu²⁺) combined with an organic binder. The photostimulable phosphor is capable of storing a fraction of the absorbed incident energy from irradiation with X-rays, ultraviolet light, electrons or protons. When later stimulated by visible or infrared radiation, it emits photostimulated luminescence (PSL), the intensity of which is proportional to the absorbed radiation energy.

An inside look

When the imaging plate absorbs incoming radiation, some of the Eu²⁺ ions are further ionized to Eu³⁺, liberating electrons to the conduction band of the phosphor crystals. The electrons, in turn, are trapped in bromine vacancies which were introduced in the phosphor crystals during the manufacturing process, forming temporary colour centres termed F-centres. Exposure to visible light releases the trapped electrons from the bromine vacancies in the lattice back to the conduc-

Table 1 Comparison of exposure time between photographic film and imaging plate method.

Method	Isotope	Photographic film	Imaging plate	Reduction rate
Colony hybridization	³² P	2 days	20 minutes	1/150
Dot blot hybridization	³² P	Cannot detect	13 hours	?
Southern blot hybridization	³² P	2 days	14 minutes	1/200
Sequencing gel	³² P	20 hours	14 minutes	1/90
Thin-layer chromatography	¹⁴ C	10 days	4 hours	1/60
Whole-body autoradiography	¹⁴ C	4 weeks	4 hours	1/170
Pulsed field gel electrophoresis	¹⁴ C	2 months	4 hours ~ 2 days	1/360 ~ 1/30

tion band of the crystal, where they go on to convert Eu³⁺ ions to excited Eu²⁺ ions. Eu²⁺ luminescence is then emitted⁶.

Because the response time of the PSL is as short as 0.8 μ s, it is possible to read between four and six megabytes of image data within several tens of seconds using a scanning laser beam. The wavelength of the PSL ($\lambda \sim 390$ nm) is reasonably separated from that of the stimulating light ($\lambda = 632.8$ nm), allowing it to be collected by a conventional high-quantum efficiency photomultiplier tube (PMT). The output of the PMT is logarithmically amplified and converted to a 10-bit-depth digital image, which can be processed by computer and displayed on a computer screen or printed as a hard copy. The residual image on the imaging plate can be erased completely by irradiation with visible light, to allow repeated use.

Advantages

The imaging plate system has several characteristic performance advantages when compared to other image sensors⁷.

First, the scanning pitch of the laser beam readout unit is 5–10 pixels per millimetre, and the full width at half maximum (FWHM) of the point spread function is $150 \mu\text{m} \times 150 \mu\text{m}$ for 10 pixels per millimetre, giving the imaging plate higher resolution than many other devices.

Secondly, compared to other integrating-type detectors, such as film and X-ray television cameras, the dynamic range of the imaging plate system is much wider, on the order of 10^5 . The response of the PSL is linear in the range from eight X-ray photons per pixel to 4×10^5 photons per pixel ($1:5 \times 10^5$), with an error rate of less than five per cent.

The detective quantum efficiency

(DQE) of the imaging plate is also better than 80 per cent for CuK β (8.9 keV) and MoK α (17.4 keV) X-rays (Fig. 1). Such a high DQE arises from both the high absorption efficiency of the phosphor for X-rays and the extremely low background noise level of the system. The noise level corresponds to less than three X-ray photons per pixel. This value compares favourably with the chemical "fog level" of film, which amounts to 1,000 X-ray photons per equivalent area.

The uniformity of response of the imaging plates varies less than 1.6 per cent over the active area. Unlike X-ray television cameras and gas-type area detectors such as multiwire proportional counters (MWPC), correction for non-uniformity of response is not required. Distortion of the image is also very small (one per cent), compared to television cameras and MWPC, and because the imaging plate is an integrating-type detector, it is free from the count-rate limitations which affect detectors operating in a pulse-counting mode.

Image data are obtained directly from the imaging plate using digital values which are easily processed by computer. With minimal precautions, the plate yields reproducible results over a long period of repeated use, unlike film, whose performance is affected by slight changes in development conditions. Fading of the stored image on the imaging plate is seldom a problem for most experiments. The fading observed after two months' exposure to radiation is 10 per cent at 0 °C, 46 per cent at 20 °C and 87 per cent at 40 °C.

The imaging plate is available not only for X-rays, but also for ultraviolet light, electrons and proton beams. Imaging plates applicable for neutron beams have been made by adding gadolinium oxide to

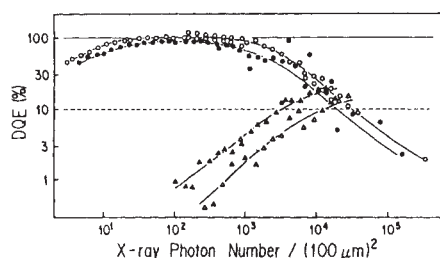


Fig. 1 Detective quantum efficiency (DQE) of the imaging plate and of film as a function of exposure level. Circles correspond to the imaging plate, and triangles to the X-ray film (Kodak DEF-5). ●, ▲: CuK β X-ray (8.9 keV); ○, △: MoK α X-ray (19.6 keV). The dashed line indicates a pulse-counting detector of 10 per cent absorption efficiency¹⁰.

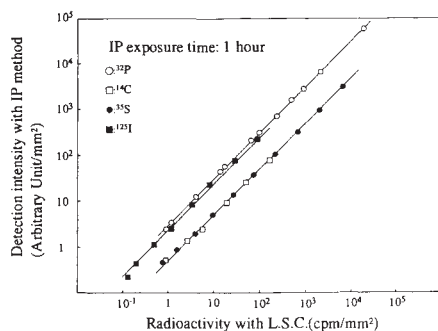


Fig. 2 Imaging plate (IP) and liquid scintillation counter (LSC) methods in the detection of various radioisotopes. The measured values for the four radioisotopes show very good correlation between the two methods, suggesting that the imaging plate has a dynamic range of five orders of magnitude.

the phosphor. The plates are easy to handle, because they are flexible, like film. Various sizes of the imaging plate are available, ranging from 250 mm × 200 mm to 430 mm × 350 mm.

Imaging plate uses

Recently, researchers in the field of X-ray diffraction have shown intense interest in the imaging plate, especially for molecular biology applications. In 1985, the usefulness of the imaging plate was demonstrated in a study of muscle contraction³. The high DQE and wide dynamic range of the imaging plate, together with its absence of count-rate limit, resulted in a sufficient reduction in exposure time to make possible the recording of a clear diffraction pattern from a contracting frog skeletal muscle in as little as ten seconds with synchrotron radiation. When compared with similar images of resting muscle, the technique allowed the molecular conformational changes during muscle contraction to be elucidated⁷.

The imaging plate has helped protein crystallographers to obtain accurate data sets by shortening the time the protein must be exposed to the X-rays, limiting the instability of the protein crystals and their damage by radiation. Using the imaging plate, a full data set can be obtained from one crystal before it is damaged by radiation^{2,8}, and with a better signal-to-noise ratio than with film. The imaging plate is also well-suited to the Laue diffraction method applied to protein crystals using synchrotron radiation⁹. A system which is similar to the imaging plate is being used with this method at the Cornell High-Energy Synchrotron Source (CHESS) in the United States^{10,11}. Data obtained by the imaging plate should have sufficient accuracy for use with the anomalous dispersion method for phase determination. The imaging plate has also been used with great success in small-angle X-ray scattering, X-ray Compton scattering, and X-ray diffraction under extremely high pressure and temperature because of its high sensitivity and wide

Japanese products

A host of products of Far East design are included in this week's review, including a range of affinity chromatography columns for HPLC, a supercritical fluid extraction system, and an X-ray diffractometer for extremely small samples.

YOKOGAWA Corporation has a new **intelligent laboratory recorder** that comes with either four, six or eight channels (*Reader Service No. 101*). Yokogawa says its model LR8100 recorder can be used for virtually any continuous writing lab recording requirement. Main menu range settings and chart speeds can be set up using either the membrane keypad or the rotary dial. A removable IC memory card allows the storage of up to three sets of configuration instructions, to save time in reconfiguring the instrument between applications. A separate memory card with storage of up to 256K is also available for later data analysis. The recorder's vacuum fluorescent display provides an analogue bar graph of each channel, a digital display of measured data with up to seven digits, and the date and time, chart speed and range settings. The instrument handles a range of inputs including a maximum 200 VDC, 12 types of thermocouples



Yokogawa's intelligent laboratory recorder.

and four types of RTDs. The recorder's auto-span feature allows it to recalibrate itself and reposition its pens if inputs exceed the preselected range.

For biochemical, immunological and membrane studies, Funakoshi Pharmaceutical offers **phosphatidylinositol-specific phospholipase C** derived from *Bacillus thuringiensis* (*Reader Service No.*

dynamic range^{9,12,13}.

In medical computed radiography, the imaging plate system is taking the place of conventional Roentgenograms¹⁴. The imaging plate not only requires a much smaller radiation dose to the patient, but also has made digital X-ray image diagnosis possible. The imaging plate system will enable the creation of a Picture Archiving and Communication System (PACS), and will allow diagnoses to be compared between isolated hospitals by the transmission of digital images (tele-radiology).

The imaging plate has also been attracting attention in the biotechnology industry — where autoradiography is commonly used to analyse gene and protein sequences — and in pharmaceutical metabolism research. The imaging plate is especially valuable because of its high sensitivity and accuracy, and because convenient, efficient measurements can be made through computer analysis. Figure 2 compares the imaging plate with liquid scintillation counter methods in the measurement of typical radioisotopes used in autoradiography. The radiation intensity in any portion of the image can be quantified with the same accuracy as the liquid scintillation method. The imaging plate has yielded quantitative data from 2-dimensional gels overnight, compared to the four weeks required in conventional autoradiography⁴, and from whole-body autoradiography in 1.5 hours, compared to 3

days for the conventional method. Table 1 outlines exposure times for several widely used techniques.

In addition to the above-mentioned applications, the imaging plate is just beginning to be used in the fields of transmission electron microscopy¹⁵, industrial non-destructive inspection and the detection of minute levels of cosmic rays or natural radiation. □

Yoshiyuki Amemiya is at the Photon Factory, National Laboratory for High-Energy Physics, Oho, Tsukuba Ibaraki 305, Japan. Junji Miyahara is at the Miyahara Development Center, Fuji Photo Film Company, Ltd, Miyahara, Kaisei-machi, Ashigarakami-gun, Kanagawa 258, Japan. For more information, fill in reader service number 100.

1. Sonoda, M., Takano, M., Miyahara, J. & Kato, H. *Radiology* **148**, 833-838 (1983).
2. Miyahara, J., Takahashi, K., Amemiya, Y., Kamiya, N. & Satow, Y. *Nucl. Instrum. Meth.* **A246**, 572-578 (1986).
3. Amemiya, Y., Wakabayashi, K., Tanaka, H., Ueno, Y. & Miyahara, J. *Science* **237**, 164-168 (1987).
4. Nakatsui, T. et al. in *Proc. Histocompatibility & Immunogenetics Conf.* (in the press).
5. Nakajima, E. et al. in *Proc. 2nd Int. ISSX Mtg* (in the press).
6. Takahashi, K., Miyahara, J. & Shibahara, Y. *J. electrochem. Soc.* **132**, 1492-1494 (1985).
7. Wakabayashi, K. et al. (in preparation).
8. Sakabe, N. *Acta crystallogr.* **A43** (supplement), C-8 (1987).
9. Amemiya, Y. et al. *Nucl. Instrum. Meth.* **A266**, 645 (1988).
10. Bilderback, D. et al. *Nucl. Instrum. Meth.* **A266**, 636 (1988).
11. Whiting, B., Owen, J. & Rubin, B. *Nucl. Instrum. Meth.* **A266**, 628-635 (1988).
12. Ito, F. et al. *Nucl. Instrum. Meth.* (in the press).
13. Shimomura, O. et al. *Nucl. Instrum. Meth.* (in the press).
14. Tatenko, Y., Iinuma, T. & Takano, M. *Computed Radiography* 45-185 (Springer-Verlag, Tokyo, 1987).
15. Mori, N., Oikawa, T., Katoh, K., Miyahara, J. & Harada, Y. *Ultramicroscopy* **25**, 195-202 (1988).
16. Amemiya, Y. et al. *Biophysics and Synchrotron Radiation* 61-72 (Springer-Verlag, London, 1987).

102). The enzyme can be used to generate either purified monomeric or dimeric forms of acetylcholinesterase from animal membranes. For the preparation of monomers with carboxymethyl residues, the membrane must first be treated with reducing and alkylating agents to break the disulfide bond forming the dimer, and then be digested with the phospholipase C. Acetylcholinesterase dimers can be created by simply exposing the membrane to the enzyme. The enzyme works by cleaving phosphatidylinositol in the membrane into diglyceride and *myo*-inositol-cyclic-1,2-phosphate. Funakoshi Pharmaceutical says the enzyme is also useful in studies of other phosphatidylinositol-anchored proteins, such as alkaline phosphatase and Thy-1. The company sells the enzyme in a package of five units; a unit is defined as the amount which catalyses the hydrolysis of 1 μmol of phosphatidylinositol per minute at 37 °C and pH 7.5.

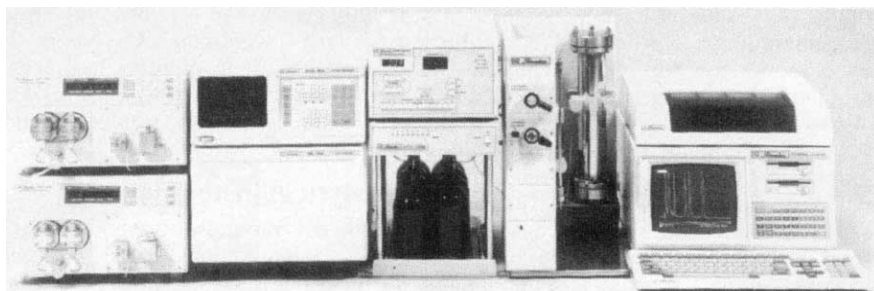
Hamamatsu Photonics sells a **dark-field observing device** with applications in nighttime observation, astronomy, microscopy and luminescence studies (*Reader Service No. 103*). The model C3100 night viewer converts low levels of incident light into electrons through its photocathode. The electronic image then passes through the night viewer's microchannel plate, where it is intensified approximately 60,000 times. The intensified image is reconverted into photons by passing through a phosphor screen, and can be viewed through the eyepiece. The optical image appears yellowish green from the phosphor screen. Hamamatsu Photonics says the model C3100 produces a high-quality picture without edge distortion. The company offers a range of accessories for the night viewer, including a zoom lens, a handgrip for ocular observation, several camera lenses and relay lenses for use in astronomy and microscopy, an infrared illuminator for use in complete darkness, and a video camera and video image processor for image analysis.

Chromatography products

Tosoh Corporation sells a range of components for **high performance gel per-**



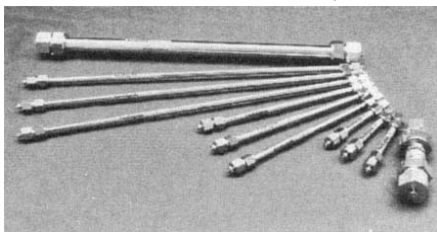
The HLC-8020 from Tosoh for GPC.



A fully configured preparative HPLC system from Shimadzu.

meation chromatography (*Reader Service No. 104*). The Tosoh model HLC-8020 gel permeation chromatograph has a parallel-flow feature which allows the stable determination of molecular weight by pumping at the sample side and reference side independently, says the company. The inlet pump is temperature-controlled, so a stable temperature can be maintained throughout the run. Tosoh says the HLC-8020's reference stop-flow provides low-pulse pumping to yield high-sensitivity analyses. The chromatograph is programmed using a membrane keypad, and a fluorescent character display shows the programme sequence and provides operating data during the run. The instrument's refractive index monitor has a Brice-type double-pass optical system for low drift and high stability. A gas detector stops the chromatograph automatically if the organic solvent should leak out. Tosoh also sells a line of columns for gel permeation chromatography under the name TSK-GEL.

For **reversed phase chromatography**, Asahi Chemical offers a column which it says combines the best properties of



Asahi's array of reversed phase columns.

octadecyl silanol gel packings with those of octadecyl polymer gels -- the Asahipak ODP column (*Reader Service No. 105*). The Asahipak ODP column packing is made by introducing C18 groups at the hydroxyl groups of vinyl alcohol copolymers. Asahi says the packing has the separation efficiency of ODS columns, but can be used with solvents of a wider range of polarities, and can be rinsed with alkaline solutions. The company's ODP-50 columns come with internal diameters of 4.6 and 6.0 mm, and are 150 mm long. The ODP-90 columns have an internal diameter of 21.5 mm, and are 300 mm long. Both have an average number of theoretical plates of over 7,000, and are functional over a pH range of 2-13. Asahi says the Asahipak ODP columns can resolve

peptides which differ by only one amino acid residue, and can separate water-soluble vitamins such as the vitamin B group. The columns are also highly stable: Asahi reports the separation efficiency of vitamin B vitamins has been maintained with an alkaline eluent for more than 1,000 consecutive analyses.

Shimadzu has put its 25 years of experience in producing liquid chromatography equipment to work in its new **intelligent preparative HPLC**, the model LC-8A (*Reader Service No. 106*). The model LC-8A is totally modular, and can be configured into a fully automatic system. Shimadzu's HPLC has a wide scale range: the flow rate of the mobile phase can be adjusted from 0.1-150 ml min⁻¹, and it operates in manual, pumped or automated modes. Using the lowest flow rate, the model LC-8A can be converted to an analytical HPLC if needed, by substituting the preparative column for an analytical column. The system's microprocessor can be programmed to alter the flow rate of the mobile phase during the run or to switch mobile phases for the most economical operation. The pump has a compression ratio control function to ensure accurate flow control of mobile phases with different physical properties. By using two or three pumps, the HPLC can be configured for gradient elution techniques. Shimadzu offers three data processors for the model LC-8A system: a menu-driven, multi-tasking programmable processor, a two-line display processor driven by pre-programmed cards, and an economical model for fundamental data processing.

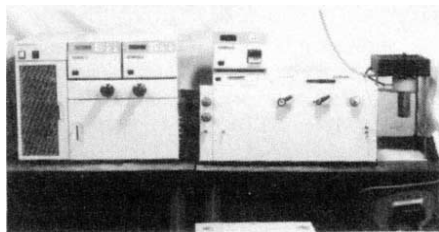
Hitachi's **intelligent autosamplers** — the models AS-3000 and AS-4000 — were developed for the integrated automation



Hitachi's autosampler has memory.

of sample dilution, sample injection, reagent addition, reaction and extraction (*Reader Service No. 107*). The autosamplers' new programmable function allows the user to compose, store and recall specific analytical and chromatographic menus, saving time in the performance of repetitive operations. The autosamplers can handle various containers, from vials and microplates to sizeable beakers, and can be used as stand-alone units or as components in a complete HPLC system. Both autosamplers have an optional reaction temperature control feature, and can be linked to a PC.

For the chemical and biotechnology industries, Jasco International Co., Ltd offers a **micro-scale supercritical fluid extraction/chromatography system** (*Reader*



Microscale SFE/SFC from Jasco.

Service No. 108). Jasco's Super-200 SFE/SFC system is adjustable for samples from a few tenths of a millilitre to tens of millilitres in volume, and comes with extraction vessels in 1, 10 and 50 ml sizes. The SFC is performed on a conventional 4.6 mm i.d. packed column. A newly designed pressure regulating system allows control of the back-pressure regardless of the mass flow-rate of the fluid, says Jasco, and dead volume is reduced to less than 10 μ l. A multiwavelength UV detector is placed between the extraction vessel and the back-pressure regulator to provide a 3-dimensional UV spectrum which can be monitored over real time during the extraction process.

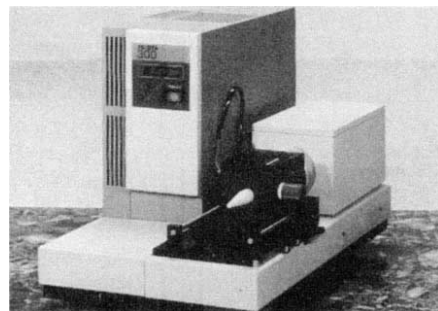
Showa Denko has expanded its Shodex range of chromatography columns to include **affinity chromatography columns for HPLC** (*Reader Service No. 109*). The Shodex AF-Pak columns are packed with a hard-type hydrophilic porous polymer 15-20 μ m gel matrix to which the ligand has been linked with a covalent bond. Ligands available include: aminobenzamidine, acriflavine, bovine serum albumin, biotin, concanavalin A, cibacron blue F3G-A, dextran sulfate, ethylenediamine-diacetic acid, N-acetyl-glucosamine, gelatin, heparin, iminodiacetic acid, ovomucoid, protein A, aminophenylboronic acid, phenylalanine, RCA 1, soybean trypsin inhibitor and WGA. The columns are packed with a buffer and antibacillus reagents suited to the ligand. Showa Denko says the columns can be used up to several hundreds of times with no ligand leaching, at pressures of as high

as 100 kgcm⁻². The columns may also be used with a common LC system. The Waters division of Millipore has recently signed an agreement with Showa Denko to market all Shodex products outside Japan.

Analytical instruments

Nozaki & Co. represents the US company Pacific Scientific in Japan. New from Pacific Scientific is the model 4300 **multi-channel particle counter**, which the company says can count and size a variety of particulate matter with diameters ranging between 0.5 and 9,000 μ m (*Reader Service No. 110*). The modular design of the model 4300 makes it adaptable for use in the laboratory or in industry, for the analysis of dry powders, liquid dispersions and aerosols. Particles are counted and sized individually using either light extinction or light scattering detection principles. Pacific Scientific says the counter yields size distribution data with a much higher resolution than that achieved by other methods, such as the mathematical analysis of Fraunhofer diffraction patterns. The counter has 32 channels and an RS232 port for interfacing with a PC.

Seiko Instruments has just come out with a new **thermal analysis system** — the Thermo-Gravimetry/Differential Thermal Analyser (*Reader Service No. 111*). Seiko Instruments' TG/DTA simultaneously



Thermal analysis from Seiko Instruments.

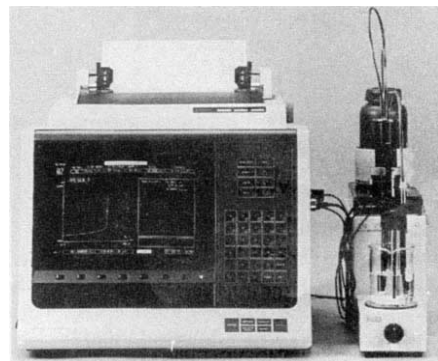
measures both thermo-gravimetry and differential thermal analysis data from a single sample to allow laboratories quickly to characterize a material on a combined thermogram. The instrument has a horizontal differential balance which is highly resistant to gas flow or convection effects, says the company, permitting gas exchange at a rate of up to 1,000 mlmin⁻¹ with no disturbance in signal, even when switching gas. Two different models are available: the TG/DTA 200 operates from ambient temperature to 1,100 °C, and the TG/DTA 300 operates at temperatures as high as 1,500 °C. Both operate at a specified TG measurement range of ± 0.1 mg to ± 200 mg and a DTA measurement range of ± 2.5 to $\pm 2,500$ μ V.

Rigaku's PSPC/MDG **X-ray microdiffractometer** has solved the problem of the X-ray diffraction analysis of extremely

small areas or extremely small samples (*Reader Service No. 112*). The company has used a curved position-sensitive proportional counter (PSPC) on a dedicated goniometer with a three axis (ω , χ , ϕ) sample oscillation mechanism to produce high detection sensitivities from a micro sample while permitting the detection of all reflections from a very wide angle of between 0 and 150 degrees 2θ . Rigaku says areas as small as 11 μ m in diameter and samples as small as 5 μ g can be measured with high accuracy by the microdiffractometer.

Measurement devices

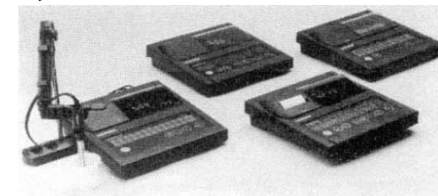
A **potentiometric automatic titrator** with a full-screen amber CRT display and a



Automated titration from Kyoto Electronics.

floppy disk drive is new from Kyoto Electronics (*Reader Service No. 113*). The model AT-310's built-in disk drive allows special titrations to be run by simply loading a disk, and data stored on the disk can be stored for re-analysis or correction. The titrator can perform up to five different titrations -- full, automatic endpoint stop, manual endpoint stop, endpoint potential/pH level, and stat titrations — sequentially by automatically switching burettes and preamplifiers. The screen allows the titration to be viewed in real time, as the titration curve develops, and the companion printer allows results to be printed out immediately.

Horiba has a versatile range of **bench-top pH meters** for use in biomedical research, quality control and clinical laboratory analysis (*Reader Service No. 114*). Horiba's F series pH meters measure and display sample temperature and pH with a 0.01 or 0.001 resolution, or millivolts with a 1 or 0.1 resolution. Automatic temperature compensation and correction capabilities enable accurate pH measurement in any fluid, says Horiba. All F series meters



Horiba's pH meters come with many features.

offer automatic calibration and automatic recognition of five standard buffers, and have an auto-hold function which provides a steady data display, and that can be overridden when necessary. Specific models in the F series also offer self-diagnosis, built-in memory of up to 100 measurements, a thermal printer, an RS-232 port for communication with a computer, and an LCD display.

The RX-1000 **digital refractometer** from Atago allows the direct reading of refractive index, temperature-corrected

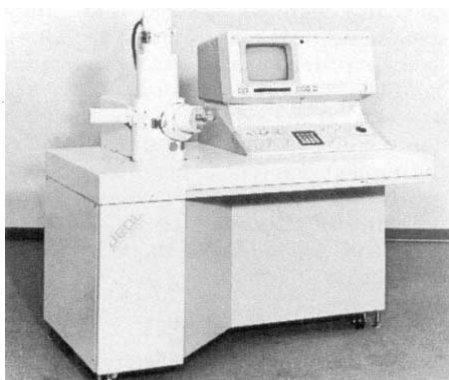


Measures refraction and temperature.

refractive index, temperature-corrected Brix per cent, and temperature-corrected concentration of various liquids (*Reader Service No. 115*). Atago's refractometer is equipped with a built-in microprocessor, and displays the critical angle of refraction detected through a photosensor on its LCD display. The temperature — detected by a thermistor embedded in the prism — is also displayed on the LCD panel. Atago says the sapphire prism is durable and resists scratching and corrosion. The RX-1000 comes with an RS-232 port for communication with a computer, and has an optional printer and constant temperature bath.

Microscopy

Jeol says its model JSM-5200 **scanning electron microscope** is a compact memory SEM that can be operated by virtually anyone (*Reader Service No. 116*). The JSM-5200's major operations are control-



Jeol's easy-to-use SEM.

led through a single keypad, with memory functions for astigmatism correction, automatic contrast, brightness control and magnification, film data input, and checking. Values for astigmatism correction are stored for all combinations of accelerating voltage and working distance, and users can program new astigmatism corrections for special applications. By pressing the "view" button, the magnification can be changed instantaneously to a pre-set magnification — a function Jeol says is especially useful for field searches. The SEM uses a low accelerating voltage, so many specimens can be observed without coating. The accelerating voltage, magnification, micron value and film number are recorded on each photograph.

Nikon says its Alphaphot 2 **classroom and laboratory microscope** offers clinical-level optics at a reasonable price (*Reader Service No. 117*). The Alphaphot 2 has a range of achromat and plan achromat objectives which use Nikon's CF system for independently correcting for chromatic aberration in both the eyepiece and the objective. Brightfield objectives are available with magnifications from $\times 4$ to $\times 100$. The Alphaphot 2's removable N.A. 1.25 Abbe condenser incorporates an aspherical lens for brighter illumination levels and an iris diaphragm for resolution and contrast control. The microscope comes with Nikon's NCB-10 type Mitsubishi rayon blue filter for correct colour rendition and a 30W tungsten lamp.

Laboratory environments

Sanyo offers a **biological clean bench** for protection against contamination (*Reader Service No. 118*). The model 13BSU has a work area that is 1,360 mm wide, 610 mm deep and 680 mm high. The clean bench has a forced-air circulation system which works at negative pressure with an air flow rate of more than 0.4 msec^{-1} , and a HEPA filter which is certified as 99.99 per cent effective. Sanyo's bench also has two UV lights along the width of the unit for added protection.

Tabai Espec has recently introduced a new range of **CO₂ incubators** designed to be lightweight and easy to use (*Reader Service No. 119*). The incubators are made of a lightweight resin moulding and have their control unit and operation panel located on the lower portion of the door and chamber, instead of on the side, to save space. The water-jacketed incubators have a low-velocity blower-circulation system that Tabai Espec says prevents contamination from condensation while greatly reducing the recovery time for temperature, humidity and CO₂ concentration after the door is opened. The incubators operate over a temperature range of 0-35 °C, and the inside corners are rounded and the humidifier pans are

removable to aid in sterilization. Conditions are set using the microprocessor keypad, which can be locked to prevent tampering or accident. An alarm sounds if the CO₂ concentration deviates from the set value by one per cent.

Yamato offers a thermo-electronically controlled **laboratory circulator** for heating and cooling needs (*Reader Service No. 120*). The Coolnics circulator has a digital temperature indicator to display its operating temperature range of 0-70 °C, and a back-up circuit for safety. Yamato says the thermoelement used in the Coolnics circulator is more energy-efficient than conventional refrigerated bath circulators, and that its PD temperature control mechanism and platinum thermoresistance element allows it to maintain a near-ambient constant temperature with an accuracy of between ± 0.01 and ± 0.03 °C. Because the circulator is air-cooled, it does not require water service, and a new noiseless circuit allows it to be used near precision equipment. □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

ADVERTISEMENTS

Medwire Div.

stocks a wide variety of

TEFLON COATED*

Platinum alloys, stainless steel, silver, multistrand and single strand available for prompt shipment. Write for descriptive brochure and price lists.



Medwire Div.

Affiliate of Sigmund Cohn Corp

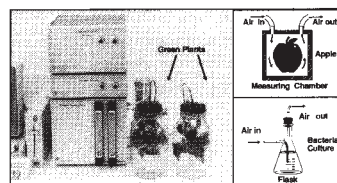
123 So. Columbus Ave.
Mt. Vernon, N.Y. 10553
914-664-5300

*Reg. T.M. of E.I. DuPont

Reader Service No.45

O₂/CO₂ micro-Respirometer

Measures gaseous O₂/CO₂ uptake or production with 0.2 µl/h accuracy
Far superior to the Warburg Apparatus



It can be used for measuring O₂/CO₂ production/consumption of insects, fish, small animals, plants, fruits, vegetables, bacterial and tissue cultures, fermentation, oxidation of metals and plastics etc. The results are printed periodically by IBM-PC/XT/AT.

Equipment is also available to measure oxygen consumption in larger animals (mice, rats, pigs, horses etc.).



COLUMBUS INSTRUMENTS

P.O. BOX 44049 COLUMBUS, OHIO 43204 U.S.A.
PHONE: (614) 488-6176 FAX: (614) 276-0529
TLX: 246514

Is teaching ready for the part-timer?

Richard Pearson

Britain's schools are struggling to come to terms with the challenge of a changing society.

THE school system is the life blood of British science, and the 1990s promise to be a decade of dramatic change as the government's education reforms start to be implemented. Will the staff be there to cope with the mixture of increased institutional autonomy, tighter curricula framework and the need to respond to market forces?

There are more than 400,000 teachers employed in state schools in England and Wales, 95 per cent of whom work full time, and although the number has been falling in recent years there is growing concern about the shortages of teachers, particularly in science-based subjects.

One of the problems that cloud debate about the solutions to teacher shortages is the lack of reliable data. At the basic level, we do not even know how many people are employed in the education system: the published total of 429,000 refers to the number of full-time equivalents (FTEs), not individuals. In addition, there are a further 50,000 teachers (FTEs) in the independent (non-state) sector. More importantly there are also a further 395,000 qualified teachers not currently working in the profession.

The precise number of vacancies is also unknown and largely unquantifiable as posts may be withdrawn, or the number of hours for a particular subject in the timetable reduced if the school thinks vacancies cannot be filled — this is known as a 'suppressed' shortage. Alternatively, pupils may be taught by teachers who are not qualified or trained in the subject, a 'hidden' shortage. Nevertheless, the available data show a vacancy rate in secondary schools of 1.1 per cent in 1987, up from 0.78 per cent in 1982. On a subject basis then, the highest vacancy rates are in computing and business and commercial studies, while the high levels in mathematics, physics and CDT (craft design technology) seemed to have eased for last year at least. By contrast, vacancy levels have been rising in languages and some other subjects. On a regional basis, Greater London suffers most at secondary level with a vacancy rate of around 3 per cent, but this conceals wide local variations from a suburban 0.4 per cent to an inner city area with a vacancy rate of 5.6 per cent.

The other side of the shortage picture is that of wastage, and the evidence here is

of a relatively stable profession with losses from teaching averaging 5.5 per cent, the dominant destination being retirement and out of employment altogether. Only one in ten of the secondary teachers leaving went to jobs outside teaching, a loss rate of less than 1 per cent from the profession.

Recruits to fill these and other vacancies come from three main sources. In the 1970s, entry to the profession was dominated by school-leavers going to teacher training college, supplemented by a cohort of graduates, some but by no means all undertaking postgraduate teaching courses. Together they accounted for nearly three out of four of those taking up teaching posts. Since that time, there have been major changes in basic teacher training, with the old teacher training colleges being absorbed by the polytechnics and colleges and the main qualification, now the Bachelor of Education, achieving degree status. For the past decade, entrants to teaching have had to undergo special training, although this requirement was initially waived in the case of shortage subjects such as mathematics, a policy which allowed a continuing inflow of untrained teachers into the profession. This may to some extent explain the comparatively low attainment level subsequently experienced in these subjects. Overall, the numbers completing initial teacher training fell from around 40,000 to under 15,000 in the decade to 1986, a decline managed by the government in the light of reducing job opportunities for teachers. Not all of those trained went straight into teaching, the figure for the early 1980s being under 60 per cent although it has since been rising to reach 65 per cent in 1986. Two thirds of these entrants were women, and 60 per cent were aged under 25.

While the traditional entry cohort of young entrants has been falling fast, the number of re-entrants initially fell from 14,000 to 11,000 but has since remained broadly stable. As a result, their share of the annual intake has risen from 1 in 4 to 1 in 2 over the past decade. These re-entrants are, as might be expected, predominantly women (90 per cent of the total) and clustered in the 30–45 age range. The third entry group are those from abroad or those without previous training and they have remained a small

(2–3 per cent) but increasing proportion.

What then of the future? The number of teachers needed depends particularly on the size of the school population, which is expected to continue to fall at secondary level for much of the next decade although the number of children at primary age will be rising. On current policies, the number of secondary teachers needed could decline by about 7,000, an overall decline of about 10 per cent. Of equal importance, however, are the effects of changes in educational policy such as the planned core curriculum, which should in theory boost the demand for teachers in subjects such as mathematics, science and languages. The level of government expenditure for education is also critical and as yet unknown. While the impact on the numbers of teachers employed must be unknown at this stage, it seems more likely that they will fall rather than rise over the next decade.

There is a fine balance between the numbers leaving the profession and those joining it, with problems in certain subjects and localities. On the supply side, organizers of teacher training will be facing increasing competition for the shrinking supply of school leavers (see *Nature* 335, 100; 1988) and of graduates to enter postgraduate teacher training (see *Nature* 334, 90; 1988). It has been estimated that next year 8.2 per cent of school leavers with two A-levels will have to enter teacher training (up from 6.2 per cent in 1984) while 1 in 10 graduates will have to do so, a figure that rises to 1 in 5 in the case of mathematics and computing graduates and 1 in 2 of those graduating in French. Given the growing demand for graduates elsewhere in the economy, a factor that may also increase wastage rates, these seem impossible targets unless the attractiveness of teaching as a career is significantly improved. A more hopeful strategy would be to increase the intake of mature entrants to teacher training and to boost the number of re-entrants from the pool of nearly 400,000 trained teachers who are not currently working. To do this, more opportunities will have to be made for part-timers. □

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.